nature

www.nature.com/nature

Vol 455 | Issue no. 7215 | 16 October 2008

More than one bad apple

A congressional investigation alleges that some researchers have failed to report all the drug-company money that they have received — and that universities may have been too slow to police them.

The case of Charles Nemeroff, who as chair of the psychiatry department at Emory University in Atlanta, Georgia, allegedly underreported his income from drug companies, offers some stark revelations. Not only does it seem that Nemeroff was able to skirt around rules for reporting income, but Emory's officials appeared unable to rein him in.

A string of internal Emory documents and e-mails made public last week after a hearing of the US Senate Committee on Finance, chaired by Senator Charles Grassley (Republican, Iowa), allege a web of consulting, lecturing and advisory-board relationships that Nemeroff maintained with 16 pharmaceutical companies. By obtaining figures from each of the companies and comparing them with Nemeroff's financial disclosure forms provided by Emory, the committee's investigators alleged that, in breach of university rules, he failed to report at least $1.2 million in income that these relationships earned him between 2000 and 2006.

In fairness, Emory medical school's Conflict of Interest Committee conducted an in-house investigation of Nemeroff's consulting in 2004. The committee alleged that he had committed "serious" violations of rules on reporting financial interests. Nemeroff accordingly promised in writing to keep his consulting with drugmaker GlaxoSmithKline (GSK) to less than $10,000 annually — the threshold beyond which, under National Institutes of Health (NIH) rules, institutions must actively manage or eliminate entirely their NIH-funded researchers' conflicts.

GSK was of special concern because, from 2003 to 2008, Nemeroff was principal investigator on a $4-million NIH grant examining five GSK antidepressant candidates. Figures given to Grassley's investigators by GSK allege that, despite his written promise, Nemeroff's compensation from the company totalled $171,000, $78,000 and $33,000 in 2004, 2005 and 2006, respectively. His disclosure forms to Emory in those same years allegedly showed payments of $9,999 per year, or of no specified amount.

Asked by *Nature* last week why Nemeroff was not disciplined in 2004 when his violations first emerged, the university responded that its 2004 internal report documenting his breach of the rules "speaks for itself". It added that the public documents show that it has "worked diligently" with Nemeroff to manage his alleged conflicts of interest. The university also noted that it is planning a "thorough investigation" of the allegations made by Grassley and that it is emphasizing its disclosure policies to all staff members.

The 2004 report did indeed lay down the law to Nemeroff. Yet Emory officials did not audit his income to be sure that he had reformed. Nemeroff stepped down as psychiatry chair only earlier this month, on the eve of the first newspaper report detailing his actions. Of course, the university's financial interests are tied up in the case too. Some $1.35 million of the NIH grant went to Emory for overhead costs. Nemeroff did not respond to *Nature*'s requests for comment. However, in a statement issued by Emory, he said: "To the best of my knowledge, I have followed the appropriate University regulations concerning financial disclosures. … I will cooperate fully and work with Emory to respond to the alleged conflicts of interest issues raised by Senator Grassley and his staff."

It is tempting to dismiss this case as a 'one-bad-apple' situation. But Nemeroff is the seventh academic psychiatrist this year that Grassley has exposed as allegedly underreporting drug-company income. His office says that there are more revelations to come. Grassley has begun pressuring the NIH to mete out real punishment — as in, pulling grants — to spur institutions to enforce proper reporting. The agency came close last week, when it imposed conditions on Emory requiring written assurance that proper disclosures have been made for every grant before it will give funds to the university. Departing NIH director, Elias Zerhouni, has also launched a time-intensive administrative revamp of the rules governing conflict-of-interest reporting by universities. His successor should make it a priority to speed up this process.

Grassley, meanwhile, has proposed a Physician Payments Sunshine Act. Introduced in Congress last year, it would legally oblige drug and device companies to post in a publicly accessible database all payments of more than $500 that they make to physicians. This would markedly ease the task of universities in enforcing proper reporting, and would exonerate the many researchers who play by the rules but who are nonetheless being cast under a cloud by Grassley's investigation. Congress ought to pass the bill into law when it next convenes. ■

Cut-throat savings

In an attempt to boost its struggling economy, Italy's government is focusing on easy, but unwise, targets.

It is a dark and angry time for scientists in Italy, faced as they are with a government acting out its own peculiar cost-cutting philosophy. Last week, tens of thousands of researchers took to the streets to register their opposition to a proposed bill designed to control civil-service spending (see page 840). If passed, as expected, the bill would dispose of nearly 2,000 temporary research staff, who are the backbone of the country's grossly understaffed research institutions — and about half of whom had already been selected for permanent jobs.

Even as the scientists were marching, Silvio Berlusconi's centre-right government, which took office in May, decreed that the budgets of both universities and research could be used as funds to shore up Italy's banks and credit institutes. This is not the first time that Berlusconi has targeted universities. In August, he signed a decree that cut

university budgets by 10% and allowed only one in five of any vacant academic positions to be filled. It also allowed universities to convert into private foundations to bring in additional income. Given the current climate, university rectors believe that the latter step will be used to justify further budget cuts, and that it will eventually compel them to drop courses that have little commercial value, such as the classics, or even basic sciences. As that bombshell hit at the beginning of the summer holidays, the implications have only just been fully recognized — too late, as the decree is now being transformed into law.

Meanwhile, the government's minister for education, universities and research, Mariastella Gelmini, has remained silent on all issues related to her ministry except secondary schools, and has allowed major and destructive governmental decisions to be carried through without raising objection. She has refused to meet with scientists and academics to hear their concerns, or explain to them the policies that seem to require their sacrifice. And she has failed to delegate an undersecretary to handle these issues in her place.

Scientific organizations affected by the civil-service bill have instead been received by the bill's designer, Renato Brunetta, minister of public administration and innovation. Brunetta maintains that little can be done to stop or change the bill — even though it is still being discussed in committees, and has yet to be voted on by both chambers. In a newspaper interview, Brunetta also likened researchers to *capitani di ventura*, or Renaissance mercenary adventurers, saying that to give them permanent jobs would be "a little like killing them". This misrepresents an issue that researchers have explained to him — that any country's scientific base requires a healthy ratio of permanent to temporary staff, with the latter (such as postdocs) circulating between solid, well equipped, permanent research labs. In Italy, scientists tried to tell Brunetta, this ratio has become very unhealthy.

The Berlusconi government may feel that draconian budget measures are necessary, but its attacks on Italy's research base are unwise and short-sighted. The government has treated research as just another expense to be cut, when in fact it is better seen as an investment in building a twenty-first-century knowledge economy. Indeed, Italy has already embraced this concept by signing up to the European Union's 2000 Lisbon agenda, in which member states pledged to raise their research and development (R&D) budgets to 3% of their gross domestic product. Italy, a G8 country, has one of the lowest R&D expenditures in that group — at barely 1.1%, less than half that of comparable countries such as France and Germany.

The government needs to consider more than short-term gains brought about through a system of decrees made easy by compliant ministers. If it wants to prepare a realistic future for Italy, as it should, it should not idly reference the distant past, but understand how research works in Europe in the present. ■

Meeting expectations

Scientists need to ask themselves if their meeting or conference is really necessary.

This week, *Nature* publishes the last in a series of essays on 'Meetings that Changed the World', with an account of a conference held in 1986 in Santa Fe, New Mexico, that helped launch the human genome project (see page 876).

The meetings highlighted in the series were unusual in that they deployed the latest science in support of larger goals. Yet most scientific meetings do not aspire to such heights. Indeed, scientists these days rarely expect to hear much new science at a conference; rather, the greatest value of meetings comes from interaction and networking. At the same time, there are now so many meetings that it is impossible for scientists to attend more than a fraction of what is on offer. So are scientific meetings really necessary?

The traditional scientific conference performs many functions. The power of face-to-face contact in generating new thinking, ideas, networks and collaborations cannot be underestimated. Moreover, increasing work and time pressures make it more important than ever to escape the daily grind and meet colleagues from around the world. Another function of scientific conferences is often to generate income for universities and learned societies, not to mention the profitable industry of conference organization.

The ever-increasing number of scientific meetings is cause for concern. Yes, the pace of science is quickening. But the proliferation of meetings is sometimes influenced as much by researchers wanting to pad out their CVs, and by the prestige conferred on an institution by hosting such an event, as it is with a desire for real intellectual exchange. All too often, meetings lack clarity of purpose and seem hastily constructed. This is particularly the case with 'me-too' conferences held to capitalize on a topical issue, such as avian flu. And the economic crisis, along with the rising costs of air travel and its impact on climate change, argues for greater parsimony and prioritization of conferences.

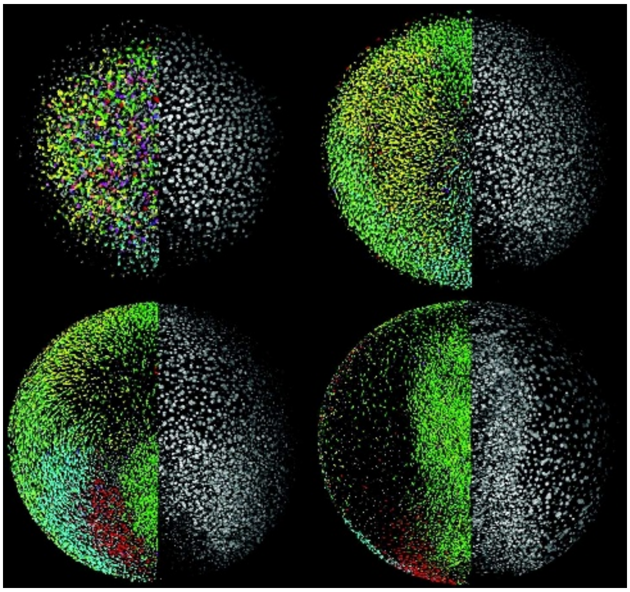
If a conference is absolutely necessary, some basic guidelines are in order. First and foremost, organizers need to be clear about a meeting's aims and objectives. Second, the number and length of formal presentations could be reduced. Attendees can now digest content before conferences begin, for example using wikis, social networks and other online tools, which leaves more time for face-to-face discussion, brainstorming and the all-important networking breaks at the event itself. Such measures would also make the content of conferences and workshops accessible to those unable to attend — particularly students, scientists from poorer countries and scientists from other, less-related fields. Third, more meetings should be webcast live, with videos archived online and linked to associated content such as papers presented, live blogging and other social networking.

Technology cannot — at least for the time being — match the power of direct interaction. Conferences are where reputations are forged — the humble poster session remains important for up-and-coming researchers to get themselves noticed and as a place for discussion. Online networking itself works better with people who know one another personally, and collaborations flow naturally from people who enjoy good relationships.

All of this means that scientific conferences are necessary and retain an important role in the research enterprise — but also that more careful thought needs to be exercised before sending out yet another call for papers. ■

RESEARCH HIGHLIGHTS

Zebrafish cell zipcodes



SCIENCE

Science doi:10.1126/science.1162493 (2008)

A new type of fluorescent microscopy has allowed biologists to reconstruct the early development of the tropical freshwater zebrafish (*Danio rerio*), a model organism used in labs around the globe.

Philipp Keller, Joachim Wittbrodt and their colleagues developed digital scanned laser-light-sheet fluorescence microscopy at the European Molecular Biology Laboratory in Heidelberg, Germany. The technique involves labelling nuclear proteins at the one-cell stage, and moving an extremely thin laser beam vertically and horizontally through specimens, which is less harmful to them than conventional techniques and can thus be used to study cells over longer time periods.

Keller's team mapped the nuclear positions and movement of every single cell in normal and mutant zebrafish embryos during the embryos' first 24 hours (pictured). They then built a model of germ layer formation, and identified a maternally defined break in symmetry — the future body axis — as the cell ball grew.

CHEMISTRY

Prion progress

Angew. Chem. Int. Edn **47,** 8215–8219 (2008)

The details of how a non-pathogenic prion protein becomes the agent of Creutzfeld–Jakob disease are obscure, but they may hinge on an 'anchor' made of sugar and lipid by which the prion attaches to a cell membrane. This anchor has been manufactured in a laboratory by Peter Seeberger of the Swiss Federal Institute of Technology in Zurich, Christian Becker at the Max Planck Institute for Molecular Physiology and their colleagues.

The researchers started with a sugar molecule containing five subunits and a lipid chain of 18 carbon atoms. Adding the amino acid cysteine allowed the synthetic chemical to react with a sulphur-containing group called a thioester on the prion protein. An 'anchor' thus attached, the prion could stick to vesicle membranes.

MATERIALS SCIENCE

Iron option

Adv. Mater. doi:10.1002/adma.200801883 (2008)

In the world of data storage, smaller is undoubtedly better. The field of single-molecule data storage has so far focused on complexes containing 12 manganese atoms. These have many unpaired electrons, the 'spins' of which make the molecules magnetic. But they lose their strong magnetism when deposited on the surface of other materials.

Roberta Sessoli at the University of Florence in Italy and her colleagues have shown that a molecular cluster based on four iron ions is more robust. Using X-rays as a probe, the team found that a layer of these molecules retains its magnetic behaviour even when sitting atop other substances. The authors believe that further research into molecules like this one could pave the way to quantum computing.

NANOTECHNOLOGY

Leveraging antibiotics

Nature Nanotech. doi:10.1038/nnano.2008.275 (2008)

The antibiotic vancomycin kills bacteria by disrupting their cell walls. To study this process, Rachel McKendry of University College London and her colleagues have designed an array of tiny cantilevers coated with molecules similar to those found in bacterial cell walls.

When their arrays were exposed to vancomycin, the antibiotic bound to the cantilevers, altered their surface properties, and caused them to bend. Measuring the bend provided a sensitive assay for antibiotic binding. The researchers also used these data to calculate how surface stress caused by antibiotic binding percolates through membranes, eventually causing cells to burst. They hope that the new method will speed the search for weapons against antibiotic-resistant 'superbugs'.

CHEMISTRY

Sugar coated

J. Am. Chem. Soc. **130,** 13465-13470 (2008)

Metal wires one atom thick are ideal for exploring quantum effects in electronics and might one day supply ultrasensitive chemical sensors. But first they must be prevented from adsorbing molecules indiscriminately from their surroundings. Jean Christophe Lacroix and his co-workers at Paris Diderot University in France believe they have succeeded in doing this by giving atomic wires a protective 'jacket'.

Lacroix and his team grew copper wires that are one atom thick at the thinnest point. When they generated these wires in a solution of tubular cyclodextrin molecules, many of the wires showed little fluctuation in conductance compared with those generated in pure water. Moreover, the wires in solution were disrupted little by an organic salt that breaks wires in pure water.

Other tests imply, but have not yet proved, that the cyclodextrins are threaded around the wires (illustrated below).

CHEMICAL BIOLOGY

Double trouble

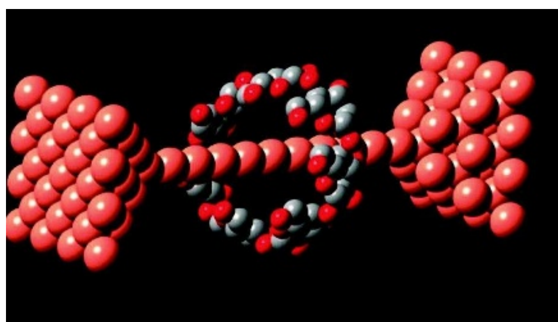
Nature Chem. Biol. doi:10.1038/nchembio.117 (2008)

'Promiscuous' chemicals that can inhibit two different classes of enzymes could aid in the design of new cancer therapies.

Kevan Shokat of the University of California, San Francisco, and his colleagues screened a library of substances, hunting for any that block both tyrosine kinases and phosphatidylinositol-3-OH kinases — enzymes that cancer drugs in development often target.

Two inhibitors fitted the bill. They belong to a category of chemicals called pyrazolopyrimidines and seem to work by

AM. CHEM. SOC.



interacting chemically with several amino acids that are conserved across the two enzyme classes. Variations in a 'gatekeeper' amino acid apparently prevented the inhibitors acting on other kinds of kinase.

MOLECULAR BIOLOGY

It's all in the timing

Cell Stem Cell **3,** 364–366 (2008)

Researchers at Mount Sinai School of Medicine in New York suggest that collecting cells for bone marrow transplantation should be done in the evening. They base this recommendation on the finding that humans typically have more blood stem cells in their circulating blood late in the day than at other times.

In February, the team, led by Paul Frenette, reported that circadian clock genes influence the movement of blood stem cells between the bone marrow and the blood in mice. Blood stem cells in humans, they now find, also show these rhythmic fluctuations, although at opposite times of day from mice.

The quotidian pattern of stem-cell levels in the blood was not disrupted in the bones of donors who were injected with a drug called AMD3100 that mobilizes blood stem cells out of the bone marrow. So, the authors reason, harvesting cells in the evening might be the best way to maximize the number collected for transplantation.

NEUROBIOLOGY

Empowering the middleman

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0806114105 (2008)

Expressing a light-sensitive protein in cells that normally mediate between the retina and the brain restores a degree of vision in blind mice.

Retinal ganglion cells transmit signals that they receive from rod and cone cells to several regions of the brain. Rods and cones are the eye's main photoreceptors, and loss of these cells — a result of a number of human diseases — causes blindness.

Ordinarily, only a small subset of retinal ganglion cells contains the photosensitive pigment melanopsin. Richard Masland of Massachusetts General Hospital in Boston and his colleagues used a virus to induce widespread melanopsin expression within these cells in mice with degenerated photoreceptors. The treatment allowed the animals to respond to light and dark.

The finding adds to previous research that manipulated another type of 'middleman' retinal cell into expressing a light-sensitive bacterial protein, channelrhodopsin-2.

GEOSCIENCES

Peatering out

Nature Geosci. doi: 10.1038/ngeo331 (2008)

If Earth were suddenly to warm by 4 °C, peatlands would lose 40% of the soil organic carbon from their shallow parts and 86% from their deep parts over about 700 years. Takeshi Ise of the Japan Agency for Marine-Earth Science and Technology in Yokohama and his colleagues make this prediction after building a model that incorporates peat chemistry, soil hydrology and, for the first time, thermal dynamics.

Peat is subject to positive feedback. The more it rains, the higher the water table, and the more oxygen-depleted the peat becomes, which lowers the decomposition rates of soil organic carbon. And the more soil organic carbon peat has, the better it retains water. Raising the temperature increases decomposition rates, so causing this feedback loop to run in reverse.

P. GOETGHELUCK

PARASITOLOGY

Hopping mad

Anim. Behav. doi:10.1016/j.anbehav.2008.07.018 (2008)

A parasitic hairworm that infects crickets and causes them to 'commit suicide' by drowning (pictured above) somehow prompts its host to develop a death wish at exactly the right time. That is, when the hairworm (*Paragordius tricuspidatus*) is sexually mature and its host's timely demise in a watery grave enables it to reproduce.

Marta Sanchez at the Centre for Evolutionary and Functional Ecology in Montpellier, France, and her colleagues have broken down the crickets' change of behaviour into two stages. They compared infected and uninfected crickets (*Nemobius sylvestris*) collected from both their native dry forest habitat and from bizarre locations such as car parks. Uninfected crickets do not stray from their natural environment; however, not all of those found in foregin territories were suicidal at the time of collection.

It seems that, as the hairworm matures, it initiates erratic behaviour in crickets, causing errant wanderings. It then triggers suicide when it is reproductively ready.

JOURNAL CLUB

Andrzej Pietrzykowski
University of Massachusetts Medical School, Worcester

A molecular biologist considers the corollary of misbehaving ion channels.

More than half a century ago, Hodgkin and Huxley hypothesized that pore-forming proteins found in a cell membrane could regulate the flow of ions across that membrane. These days, we classify ion channels according to the ions they allow through and the nature of the pore-forming protein. The crucial part of a pore is the protein's alpha subunit, which lines the pore. Auxiliary subunits, denoted by other letters of the Greek alphabet, merely tweak a channel's characteristics.

The basics infer an assumption: that different channels can interact with each other, but that subunits buried within a channel are 'married' to that channel 'for life'. A voltage-activated calcium channel can, for instance, form a pair with a large-conductance calcium-activated potassium channel. But a beta subunit of the calcium channel can associate only with the calcium channel's main alpha subunit, and a beta subunit of the potassium channel remains 'faithful' to the alpha subunit that surrounds the potassium pore.

However, assumptions should always be tested. In this case, Shengwei Zou and his colleagues at the University of Houston in Texas have taken the potassium channel in this example and shown that it is bound by an auxiliary beta-1 subunit of an L-type calcium channel ($Ca_v\beta1$). When this subunit interacts with the potassium pore, it alters both the pore's kinetics and calcium sensitivity (S. Zou *et al. Mol. Pharmacol.* **73,** 369–378; 2008).

I view this finding as part of an emerging theme, the ramifications of which could be profound. Ion channels may, in general, be much more dynamic structures than is currently recognized. This means that when researchers monitor a channel's activity they may not be recording exactly what they think they are — and that targeting ion channels with new drugs could produce unexpected side effects.

Discuss this paper at **http://blogs.nature.com/nature/journalclub**

NEWS

Mars missions face cost crunch

In Europe and America, future missions are in doubt.

Europe's next Mars mission launch looks likely to slip from 2013 to 2016. And budget overruns on NASA's next Mars rover threaten future missions there and elsewhere.

The European Space Agency (ESA) is having problems paying for an ambitious redesign of the science suite of its planned ExoMars lander, which consists of a rover with a drill and a stationary observational platform. The changes were decided in spring 2007 by ESA's programme board for human space flight, microgravity and exploration, raising the mission's estimated cost from €650 million (US$890 million) to between €1 billion and €1.2 billion. That cost hike must be approved by participating nations next month at an ESA ministerial meeting.

Italy, which contributes nearly 40% of the cost of ExoMars, had originally supported the increase. But following last April's election, the new government says it will not give any additional money. Germany is also resisting the price hike. "The solution in my mind can only be to have a redesign of the overall mission," says Johann-Dietrich Wörner, chairman of the DLR, Germany's space agency.

The immediate cash shortfall is nearly €100 million. If more money can't be found within a month, the mission will probably not launch in 2013, which could increase its costs by at least 10% and might derail it entirely. "This is make-or-break," says David Southwood, ESA's science director.

Southwood hopes that Britain, which at 15% is among the largest contributors to ExoMars, will help make up the shortfall created by Italy. But David Williams, director-general of the British National Space Centre, which oversees UK involvement in space missions, says "I don't think there's much chance of that". France has said it will increase its own percentage but not make up for other countries' deficits.

"This is make-or-break."

The trouble for ExoMars comes at the same time that NASA's Mars Science Laboratory (MSL) is facing its own financial difficulties. On 10 October, MSL officials met NASA administrator Michael Griffin to review the mission's progress, and decided to push ahead with a 2009 launch despite budget problems. Estimated in 2006 to cost $1.6 billion, the price now looks likely to be more than $2 billion.

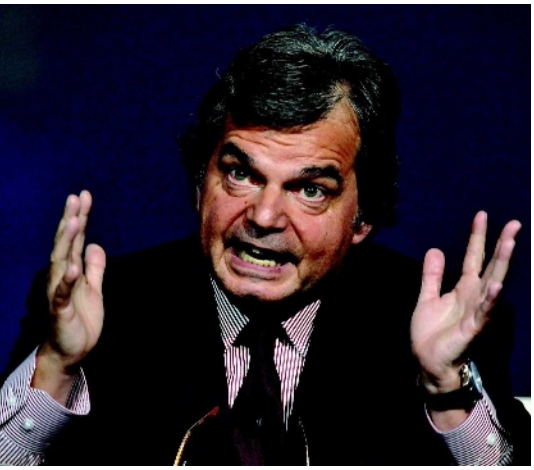
Engineers at the Jet Propulsion Laboratory (JPL) in Pasadena, California, are building the rover, and part of the cost problems lie in the new technologies being designed for the MSL, says Doug McCuistion, director of NASA's Mars exploration programme. These include actuators, the motors that drive the spacecraft's wheels and robotic arm. Ed Weiler, the agency's associate administrator for space science, says that MSL engineers did not accurately estimate how much time and money would be needed to develop these and other elements of the rover. Staff overtime is also running over budget. Some $1.5 billion has already been spent on the MSL.

Weiler says NASA will first look at the timelines for other missions under development, starting with other Mars missions and then missions to other planets. The latter could include the Jupiter mission Juno and the lunar orbiters GRAIL and LADEE. For now, principal investigators for the other missions have been told to proceed with mission planning as scheduled, says Maria Zuber of the Massachusetts Institute of Technology in Cambridge, who leads the GRAIL team.

Jack Mustard, a planetary scientist at Brown University in Providence, Rhode Island, says many feel that JPL hasn't been able to prove its ability to predict costs accurately. Giving them more money, he says, "is like enabling your drunken cousin".

JPL had asked for and received an additional $200 million this spring, and is now expected to need about $100 million more in 2009. "There are only so many times that this institution can run into these problems and expect forgiveness," says Jim Green, director

New law threatens Italian research jobs



A. CASASOLI/CONTRASTO/EYEVINE

Renato Brunetta has lambasted the civil service.

Nearly 2,000 Italian researchers will lose promised permanent positions under a law that is expected to come into force by the end of the year. They may have to leave public research altogether.

Last week, the chamber of deputies in Silvio Berlusconi's new centre-right government considered the bill, which is intended to reduce public spending by streamlining the civil service. Several researchers have put themselves up for sale on eBay, as part of a campaign that has also involved tens of thousands of protesters marching through the streets of Rome and other cities.

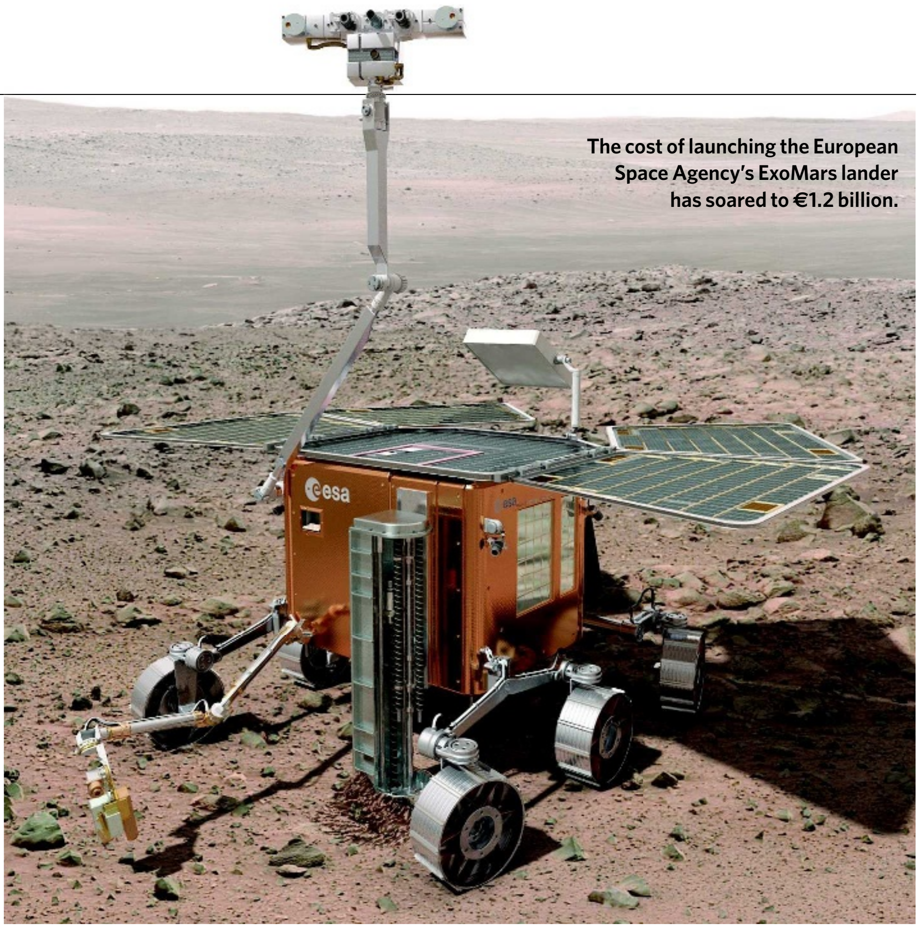
The proposed law explicitly reverses another law passed by the previous, centre-left government, whereby long-term temporary research employees could be taken on permanently if they were suitably qualified. As it also prevents scientists from being employed on sequential short-term contracts, those already selected for permanency who have had more than three years of contract in the past five years will now be let go.

Renato Brunetta, the minister for public administration and innovation who designed the new law, has enraged scientists further by referring to many civil servants as *fannulloni*, or 'idlers'.

Researchers in Italy are civil servants, and the number of positions available is determined by the central government rather than by individual research agencies. The past decade has seen almost

DIUS

INVEST FOR SUCCESS
New UK science minister pledges to defend budget in face of financial meltdown.
www.nature.com/news



ESA

The cost of launching the European Space Agency's ExoMars lander has soared to €1.2 billion.

of NASA's planetary science division.

NASA administrators and MSL personnel will meet again in early January to review the mission's progress. If the MSL doesn't look as if it will be ready to launch in 2009, it may be delayed until 2011.

The fate of other Mars missions, including the MAVEN orbiter slated for a 2013 launch and a lander planned for 2016, remains up in the air. And the idea of scooping up a sample of Mars soil and returning it to Earth might be pushed back until as late as 2022. ■

Geoff Brumfiel and Ashley Yeager, with additional reporting by Eric Hand

no new recruitment, and the number of temporary research staff has consequently rocketed. There are at least 4,500 long-term temporary staff — known as *precari*, in reference to their precarious positions — who stumble from one short-term contract to another.

The scientists say that their protest is not directed against the conventional postdoc system, but against the unhealthy ratio of temporary to permanent staff. "We have pathological numbers because new long-term positions have been blocked," says Luciano Maiani, president of the CNR, Italy's national research council.

As a result of the protests, Brunetta says that researchers will be given until 1 July 2009 while he investigates their claims. But presidents of the various Italian research agencies believe that the only way out of the situation is for the agencies to have more autonomy from the civil service.

"The government should recognize the highly specific professionalism of the research personnel — it is not appropriate for them to fall under civil-service rules," says Enzo Boschi, president of Italy's National Institute of Geophysics and Volcanology.

Claudio Gatti is a particle physicist at the National Institute of Nuclear Physics in Frascati who stands to lose a promised permanent job under the proposed law. He says that "in the Italian research system there's no planning, no mobility, no future — but we are ready to fight for our rights with every legal means available to us".

Research and education minister Mariastella Gelmini has not commented publicly on the situation, and did not respond to requests from *Nature* for comments. ■

Emiliano Feresin and Alison Abbott

See Editorial, page 835.

No more third time lucky

The US National Institutes of Health (NIH) announced last week that biomedical researchers will be able to amend and resubmit a failed funding application only once. Applicants whose grants are unfunded after the second submission may reapply only after designing a new proposal.

The new guidelines, effective from 25 January 2009, are part of an NIH overhaul of the peer-review system for evaluating grant proposals. That system previously allowed applicants two chances to resubmit rejected proposals. Earlier this year, the agency mooted doing away with resubmissions, but decided against it after an outcry from researchers (see *Nature* 453, 835; 2008).

The NIH estimates that the move will reduce the number of applications by up to 5,000 — welcome news as it struggles to evaluate about 55,000 applications this year.

Grants are increasingly awarded only after they have been through several rounds of submission — in 2007, only about 30% of awards were granted to first-time submissions — and some think resubmission has directed funding towards less competitive proposals. "The study sections may feel 'we've tortured this person long enough'," and fund them out of sympathy, says Keith Yamamoto, a molecular biologist at the University of California, San Francisco, who co-chaired a panel tasked with evaluating the NIH's peer-review system.

Toni Scarpa, director of the NIH Center for Scientific Review, says the new policy will remove delays to funding the most worthy projects, and calls it "a moral imperative".

But some researchers object to the change. Gregory Petsko, a biochemist at Brandeis University in Waltham, Massachusetts, says the system discriminates against young investigators who may need more guidance with their applications. Others view the resubmission process as a way for equally meritorious applications to wait their turn for funding in a time of tight NIH budgets.

The new guidelines could stifle worthy projects, says Beatrice Hahn, an HIV researcher at the University of Alabama, Birmingham. "What are we supposed to work on thereafter?" she says. "Although well intentioned, this change will cause major problems in the current funding crisis." ■

Heidi Ledford with additional reporting by Erika Check Hayden

Icelandic biotech feels the pinch

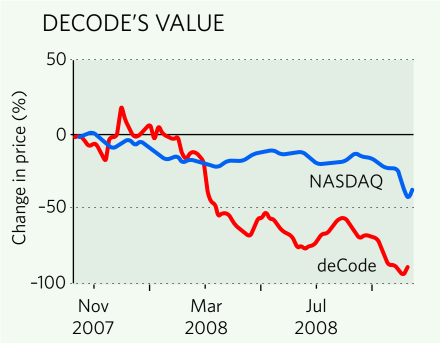
It is not surprising to find an Icelandic business doing badly these days; the country has arguably been hit the worst of any in the current financial crisis. But the serious problems being felt at Iceland's deCode Genetics come more from US markets — from investment decisions made by its bankers, it says, and from a stock price that has plunged 54% since the beginning of September.

Looked at one way, the past month has been a good one for the firm. It identified three new schizophrenia-associated mutations (H. Stefansson *et al. Nature* **455,** 232–236; 2008). It filed an application with the US Food and Drug Administration to begin human tests of a new drug for Alzheimer's disease. And earlier this week it reported in *Nature Genetics* the discovery of genetic variants associated with a risk of the cancer basal-cell carcinoma (S. N. Stacey *et al. Nature Genet.* advance online publication doi:10.1038/ng.234; 2008).

That latest announcement sent deCode's stock soaring — but only relatively so. As *Nature* went to press, it was trading at 53 cents a share. That puts its market capitalization (the number of shares multiplied by the price per share) at $32.8 million, which is below the $50 million minimum allowable by the NASDAQ stock exchange. If the company doesn't regain this minimum for a period of ten consecutive business days before 30 October, it will be delisted and lose access to NASDAQ investors.

"It's important that it keeps the NASDAQ listing," says Robin Campbell, a biotechnology-equities analyst at the London office of Jefferies International investment bank. "It's an important part of the well-being of the company from an investor and money-raising point of view."

Kári Stefánsson, the former Harvard professor who is founder and chief executive of the Reykjavik-based firm, attributes part of deCode's current travails not to the banking crisis devastating his native Iceland — that, he says, has driven down the company's local costs because of the currency devaluation — but to poor investments in auction rate securities by the US banks managing its money. "We have been seriously influenced by the American banking crisis and not the Icelandic," he says. "But we will survive. We are working on refinancing and restructuring the company. And we are fairly optimistic today that we will succeed."



"We have been seriously influenced by the American banking crisis and not the Icelandic." — Kári Stefánsson

Founded in 1996, the pioneering company drew early attention when the drug giant Roche provided $200 million to finance deCode's then-incipient gene-discovery efforts. More buzz was generated in the summer of 2000, when deCode went public at $18 a share and even spiked briefly at more than $31.

But being a frontrunner has exposed deCode to the growing pains of a field that isn't delivering as quickly as was once expected. Early in 2006, the firm's share price began a long, steady decline. And the trend wasn't helped a few months later when a drug to prevent heart attacks ran into manufacturing trouble and the company had to suspend its late-stage clinical trial. Although the firm's personal-genomics division caught media attention when it was launched in November 2007, therapeutics and diagnostics are where investors are looking for the promise of serious returns.

Survival instinct

But analysts contend that, without more specifics about the firm's current restructuring efforts — which Stefánsson says he can't give under US Securities and Exchange Commission rules — it is hard to judge the company's prospects for survival. More details may emerge during the company's quarterly conference call with investors, which will occur in the next few weeks.

In assessing deCode's chances of survival, says Campbell, three considerations will be pivotal. Can the firm find a pharmaceutical-company partner to take forward the heart attack and thrombosis drugs that have stalled in clinical trials because deCode has all but run out of cash? Can it expand its diagnostics business quickly and significantly enough to regain investor confidence? And can it find the financial wherewithal to cope with impending debt repayment obligations of $238 million? If not, says Campbell, "it might find itself between a rock and a hard place".

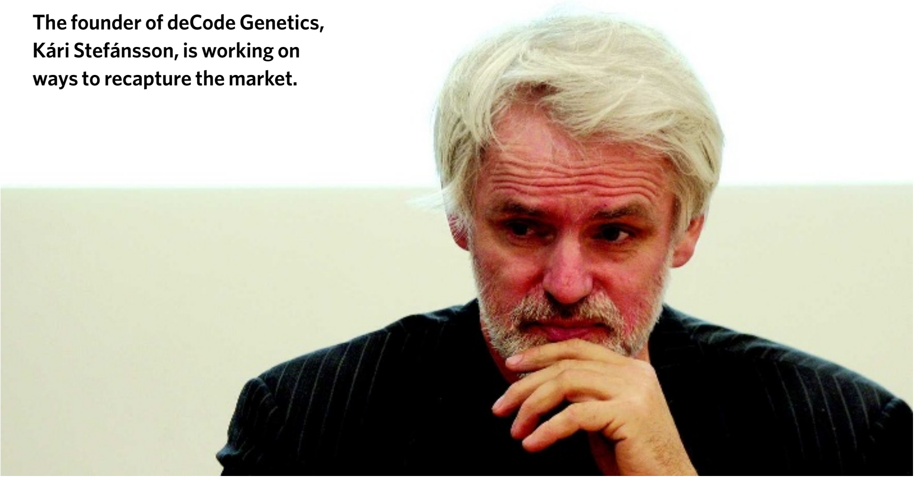
Fatal or not, the firm's situation isn't unique. "There are a lot of biotechnology companies out there right now that are feeling pain and pressure," says David Mott, a general partner in the Chevy Chase office of New Enterprise Associates, a venture-capital firm in Maryland. Mott, a former chief executive of the biotechnology firm MedImmune, points out that MedImmune went from $50 million in market capitalization to $15 billion in a five-year-period. "A fundamental advance of the business can rekindle things," he says.

For his part, Stefánsson argues that deCode is being dramatically undervalued by investors. "The infrastucture of the company has never been stronger and we have never been more productive," he says.

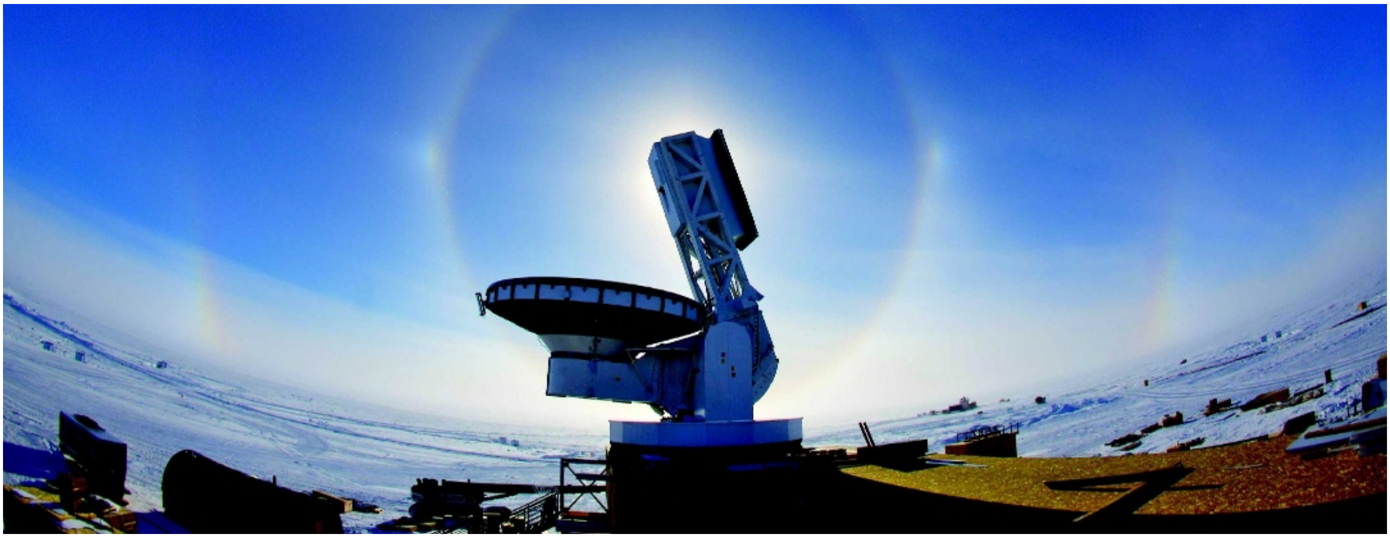
Meredith Wadman

For more on the impact of the financial crisis on science, see www.nature.com/financialcrisis.

I. LANGSDON/REUTERS


The founder of deCode Genetics, Kári Stefánsson, is working on ways to recapture the market.

Dark energy: the quest for galaxies



J. MCMAHON

By locating new galaxies, the South Pole Telescope should allow researchers to make a more accurate estimate of the qualities of dark energy.

Astronomers searching for evidence of the mysterious energy that is speeding up the expansion of the Universe have discovered three new galaxy clusters. They used a microwave survey technique that could rival existing ways of searching for dark energy.

The discovery is the first step towards a catalogue of thousands of galaxy clusters, whose evolution in the early Universe reflects the tug of war between gravity and dark energy, the repulsive force that seems to make up three-quarters of the total mass-energy in the Universe. The astronomers found the three clusters by searching for shadows in the cosmic microwave background, the relic radiation of the Big Bang, using a microwave telescope located at the Amundsen-Scott South Pole Station in Antarctica (see 'The South Pole Telescope').

"We're looking at how many clusters there are as a function of time, and that depends on the expansion factor, which leads us back to the dark energy," says John Ruhl, a collaborator from Case Western Reserve University in Cleveland, Ohio, one of seven universities and labs involved in the US$19-million telescope.

Previous dark-energy estimates relied on observations of very old and distant star explosions. Astronomers think that supernovae of a specific class explode with uniform brightness, so the observed brightness can be correlated with distance. More distant supernovae seem to be receding more quickly than they should be, given a constant expansion of the Universe, so astronomers added a fudge factor: the accelerating effect of dark energy.

The galaxy-cluster method will provide an independent check from a different direction: the growth of structure in the Universe, as hundreds of galaxies coalesce into gravitationally bound clusters, and then stop coalescing. "They'd better both give the same answers or something is wrong," says John Carlstrom, director of the Kavli Institute for Cosmological Physics at the University of Chicago in Illinois and principal investigator for the project.

He announced the discovery on 9 October at a dark-energy conference in Munich, Germany. The findings have also been published on the arXiv preprint server (Z. Staniszewski *et al.* http://arxiv.org/abs/0810.1578; 2008).

The telescope came into use in 2007 and will complete its second season of observations in November. By the end of four winters, the team expects to have a catalogue of several thousand galaxy clusters in the southern sky. That would be enough to constrain dark energy to a precision that could rival that of the Joint Dark Energy Mission, a space telescope planned for 2015 that is likely to push the 'supernovae method' to its limits (see *Nature* 455, 577; 2008).

Having two viable methods "is very important", says Mario Livio, a theorist at the Space Telescope Science Institute in Baltimore, Maryland. "The nature of dark energy is arguably the biggest question physics is facing today. You absolutely want to have more than one method to get at it." ■

Eric Hand

The South Pole Telescope

The 10-metre South Pole Telescope sits atop a plateau several kilometres high. Few clouds and little atmospheric water vapour mean that the telescope has a clear shot at identifying tiny variations in the cosmic microwave background (CMB). It relies on an array of about 1,000 tiny detectors that, cooled to a quarter of a degree above absolute zero, can determine temperature differences of several thousandths of a degree.

The CMB radiation that the telescope looks for acts as a sort of cosmic backlight. As microwaves pass through galaxy clusters, the photons are excited by hot electrons in plasma in the intergalactic space. The result is a tiny shift in the intensity of the CMB that can be seen like a shadow.

The telescope is one of the first microwave telescopes to be able to do large-sky surveys with precision. And it should be able to provide better information about the size distribution of galaxy clusters, including ones that are not yet known. The resulting three-dimensional size-distribution pattern can be used in conjunction with information from other telescopes to work out the strength of dark energy. **E.H.**

GSK

Q&A: PHARMA'S FUTURE
GSK research leaders answer *Nature*'s questions.
www.nature.com/news

Outcry at scale of inheritance project

The US National Institutes of Health (NIH) handed out the first payments in a multimillion-dollar project to explore epigenomics last month. But some researchers are voicing concerns about the scientific and economic justification for this latest 'big biology' venture.

Epigenetics, described as "inheritance, but not as we know it"[1], is now a blisteringly hot field. It is concerned with changes in gene expression that are typically inherited, but not caused by changes in gene sequence. In theory, epigenetic studies can help explain how the millions of cells in the human body can carry identical DNA but form completely different cell types, and perhaps why certain cells are susceptible to disease.

The NIH's epigenomics initiative is a plan for such studies on a grand scale including, for example, surveys in different human cell types of all the chemical tags, or epigenetic marks, that might control genes. The agency has now announced the recipients of the first US$18 million of funding. The project will eventually command $190 million over five years as part of the agency's 'Roadmap' programme.

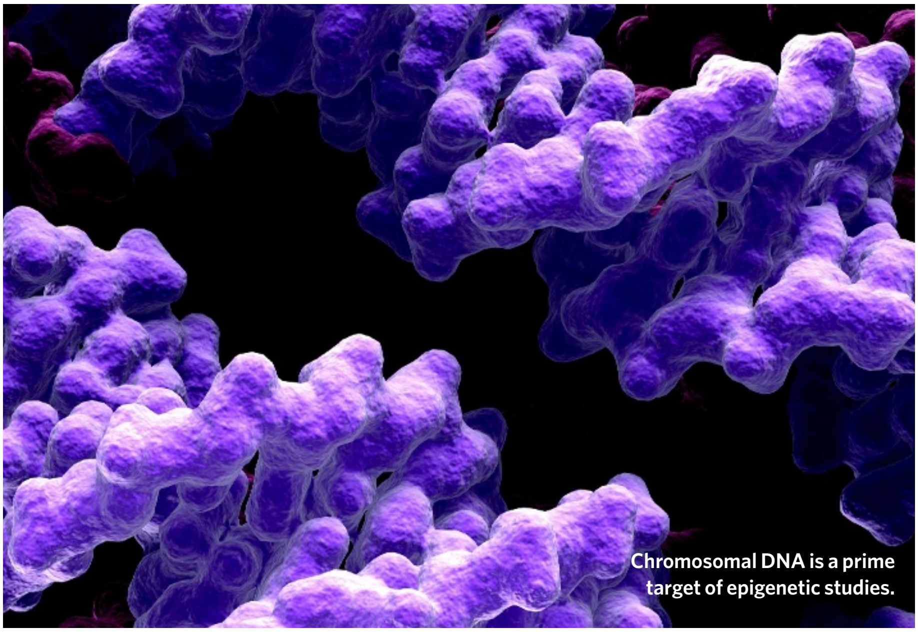
The announcement has amplified rumbling criticism of the initiative. Eight prominent scientists in the field laid out their concerns in a letter[2] published in *Science* on 3 October. Molecular biologist Hiten Madhani, the letter's lead signatory, has also set up an online petition to encourage others to register their opposition. "I hope the next NIH director just eliminates the programme lock, stock and barrel," says Madhani, who works at the University of California, San Francisco.

"We think it's a humongous waste of money for a lot of reasons," says Kevin Struhl of Harvard Medical School in Boston, Massachusetts, who is one of the letter's signatories.

Gene switches

As the emphasis in biology is switching away from genetic sequence and towards the mechanisms by which gene activity is controlled, epigenetics is becoming increasingly popular. Many studies in the field focus on chemical alterations that can switch genes on and off, including modifications to the histone proteins with which DNA is packaged into chromatin.

As part of the NIH initiative, various groups will produce 'reference' epigenomes — maps in which the genetic sequence is overlaid with chemical tags in stem cells, embryonic cell lineages and other cell types. The project will also fund efforts to find new gene-regulating marks, as well as technologies that can identify these marks in individual cells, and projects to find links between such marks and disease.



Chromosomal DNA is a prime target of epigenetic studies.

K. EWARD/BIOGRAFX/SPL

But Madhani and others say that focusing on these types of modifications — which are found all over the genome — overlooks key processes, such as the work of transcription factors and other DNA-binding proteins, that determine where these marks go. A perfect illustration of how transcription factors trump epigenetic marks, they say, is recent work by researchers in Japan showing that just four proteins can reprogram an adult cell and transform it into an embryonic-like state[3]. The problem, says Struhl, is that "transcription is passé and epigenetics is cool. It's not that epigenetics is totally useless. I just don't see why it's worth $190 million."

"It's not that epigenetics is totally useless, I just don't see why it's worth $190 million."

Some researchers in the field roll their eyes at these criticisms and say that this group and others have harped on for too long about the neglect of transcription factors. At meetings, "there is always a shouting match across the floor", says Peter Fraser, head of the chromatin laboratory at the Babraham Institute in Cambridge, UK. "It's going to be a combination of things that explain gene expression. Trying to integrate this stuff is what's important. When funding is tight," he says, "people squeal."

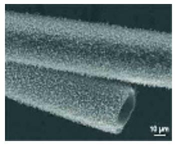
Alan Krensky, director of the NIH's Office of Portfolio Analysis and Strategic Initiatives, who is one of those leading the epigenomics initiative, says that the scientific community was consulted extensively over the programme and that a key aspect will be to integrate the information into ongoing research on transcription factors and other control mechanisms covered by different NIH grants. "This is not either/or by any means," he says. "Everything they're in favour of is part of the big picture."

The debate has implications beyond the United States — other countries are already running epigenomic projects. Earlier this year some researchers, led by Peter Jones of the USC/Norris Comprehensive Cancer Center in Los Angeles, California, called for the research community to coordinate an international human epigenome project[4] — something Madhani calls "an ill-conceived, wasteful boondoggle".

Some of these issues might be thrashed out in December at a small meeting on epigenetic mechanisms to be held at Cold Spring Harbor Laboratory's Banbury Center in New York. But the discussion is likely to be heated. Mark Ptashne of the Memorial Sloane-Kettering Cancer Center in New York, warns: "My topic at Banbury is: can anyone stand up and defend an epigenome project?" ■

Helen Pearson

1. Bird, A. *Nature* **447,** 396–398 (2007).
2. Madhani, H. D. *et al. Science* **322,** 43–44 (2008).
3. Takahashi, K. & Yamanaka, S. *Cell* **126,** 663–676 (2006).
4. The American Association for Cancer Research Human Epigenome Task Force and the European Union, Network of Excellence, Scientific Advisory Board *Nature* **454,** 711–715 (2008).



STRONG STUFF
Colossal carbon tubes take the strain.
www.nature.com/news

PHYS. REV. LETT./ H. PENG ET AL.

The death of microarrays?

High-throughput gene sequencing seems to be stealing a march on microarrays. **Heidi Ledford** looks at a genome technology facing intense competition.

Faster, cheaper DNA sequencing technology is revolutionizing the burgeoning field of personal genomics. But it is having another, more subtle effect. Scientists are using the sequencers to tackle a wide range of research applications, including monitoring gene expression, mapping where proteins bind to the genome and cataloguing sites at which the DNA has been chemically modified. In doing so, they have enabled the technology to encroach on markets that have been dominated — and created — by microarray manufacturers.

The US$700-million market for expression arrays was becoming increasingly competitive even before high-throughput sequencing entered the scene. Once controlled entirely by Affymetrix, an array maker based in Santa Clara, California, the field is now populated by several companies jostling anxiously for position. Researchers initially flocked to microarrays to measure gene expression, then to assay the human genome for the presence of specific sequence variations.

Sequencing is still more expensive than arrays and many labs do not have access to sequencing equipment, which also requires a significant amount of computing infrastructure. But prices are expected to plummet, and the new sequencing platforms are already making their way out of specialized genome sequencing centres and into academic research labs. Californian firms Illumina in San Diego and Applied Biosystems in Foster City, as well as Connecticut-based 454 Life Sciences in Branford (now owned by Roche) all report that most new orders for their machines are coming from places other than sequencing centres.

Many researchers feel that sequencing sidesteps several technical limitations of microarrays, such as their failure to detect poorly expressed genes. And mounting dissatisfaction with results from genome-wide scans for genetic links to disease — a major market for genotyping arrays — has fuelled a call for sequencing targeted regions of the genome rather than full genotyping of study participants.

Some researchers predict that array makers who specialize in tailor-made, small-market arrays to suit more specific research needs — such as Roche NimbleGen based in Madison, Wisconsin — are best positioned to cater to the future microarray market. And the low cost and quick analysis of microarrays will ensure that the technology survives for some applications, especially genotyping, says biostatistician Xiaole Shirley Liu of the Harvard School of Public Health in Boston, Massachusetts. But the fact remains: "When possible, biologists are willing to spend more to get better data," says Liu.

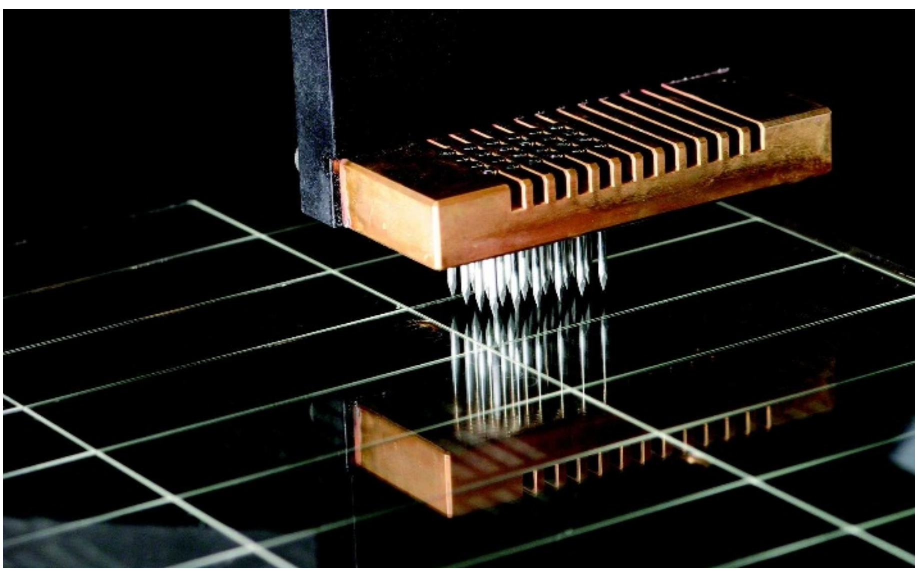
> **"Sequencing does put microarrays at risk in some areas of the life-sciences research market."**

In the past two financial quarters, Affymetrix has acknowledged the growing threat from sequencing. On 13 October, the company said that its revenues for the third quarter of 2008 would come in well below analysts' estimates. "We are very deep into a transition phase of our business," said Affymetrix president Kevin King at the UBS Global Life Sciences Conference in New York last month. "It's going to take several quarters for us to restore the growth that we have had historically."

The company blames the declining growth — revenue was down 1.5% last quarter compared with the same time last year — on shrinking budgets in the pharmaceutical industry. But some analysts think that intrusion of sequencing into the gene-expression market could be having an effect. Although most of Affymetrix's growth is in the genotyping market, sales of expression arrays still account for almost 60% of the company's consumable sales revenue, says Douglas Schenkel, a vice-president at the Boston office of investment banking firm Cowen and Company.

Affymetrix and other microarray manufacturers are striking back with improvements designed to drive prices down and expand into clinical diagnostics, a growing field best suited for mature technologies and less likely to adopt sequencing in the near future. "Sequencing does put microarrays at risk in some areas of the life-sciences research market," says John Sullivan, a research analyst at Leerink Swann, a Boston-based health-care investment bank. "But improved genetic understanding of disease actually opens up the microarray platform to clinical diagnostics."

Manufacturers are also creating arrays that can be used to isolate specific regions of the genome for sequencing, so that researchers will not have to waste resources sequencing random fragments from the entire genome. These products, called 'capture arrays', have created a new market that, notably, is tied to the success of sequencing, says Kevin Meldrum, director of business development and genomics at Agilent, an array-maker in Santa Clara. It was a logical move, he says: "If we didn't do anything, we could be sitting here five to six years from now scratching our heads." ■

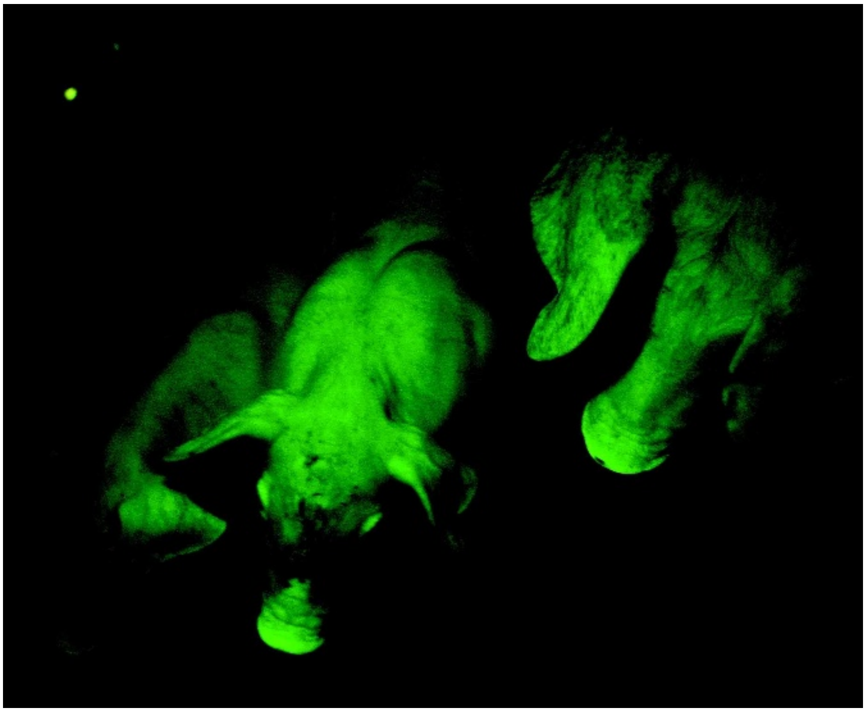


P. DUMAS/EURELIOS/SPL

DNA is deposited on a glass slide to make a microarray — but is the technology losing its allure?

SNAPSHOT

Green ham (no eggs)



NAT. TAIWAN UNIV./REUTERS

Green fluorescent protein (GFP) makes these transgenic pigs glow eerily, but it has also netted its discoverers this year's Nobel Prize in Chemistry.

The prize is shared between Osamu Shimomura from the Marine Biological Laboratory in Woods Hole, Massachusetts; Martin Chalfie of Columbia University in New York; and Roger Tsien at the University of California, San Diego.

In the 1960s, Shimomura isolated the protein from some *Aequorea victoria* jellyfish he had collected. Thirty years later Chalfie showed that the gene for the glow could be expressed by other organisms, allowing genetic processes to be traced in real time. The effects can be dramatic: these glowing pigs have had the jellyfish gene added onto their genes.

Tsien took the field further by tuning GFP chemistry to create similar proteins in other colours, so that different proteins can be seen interacting, or so that neurons firing in the brains of mice light up in multicolour.

To read an interview with Chalfie, see http://tinyurl.com/3u6vwv.

Pakistan's education head is asked to resign

Pakistan's higher-education minister has stepped down after a change in government.

Atta-ur-Rahman, an organic chemist at the University of Karachi, had served as head of the nation's Higher Education Commission since 2002, but his position became uncertain after a national election last month. Last week Pakistan's new president, Asif Ali Zardari, asked for Rahman's resignation. According to Rahman, his departure comes in the midst of a financial crisis that has seen the higher-education system's 27-billion-rupee (US$340-million) budget cut by one-fifth.

Rahman is widely credited with raising the salaries of professors and boosting research spending. "A lot of good has been done," he says. "I hope the person who succeeds me will be able to continue it going forwards."

Europe plans fewer free emissions allowances

Europe's ambitious climate plan was approved last week by the European Parliament's environment committee. But a group of eastern European countries, led by Poland, threaten to block reforms to the emissions-trading system, saying that these would harm their economies.

The package of market-based and regulatory measures is intended to help reduce greenhouse-gas emissions across the European Union by 20% (compared with 1990 levels) by 2020 (see *Nature* 451, 504–505; 2008). The target will be increased to 30% when a new international climate-change agreement is reached.

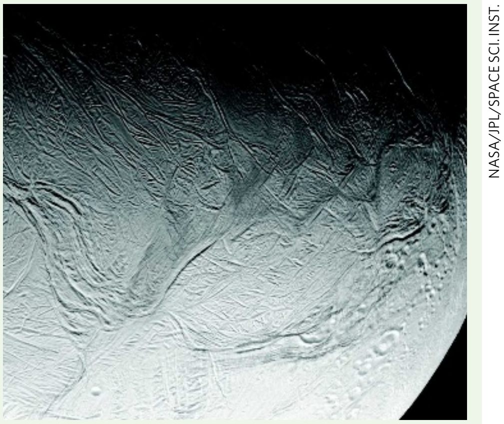
The committee backed a number of contested changes. From 2013, the power sector will have to acquire all of its emissions allowances at auction, instead of getting them for free. Large manufacturing industries that compete internationally will be gradually phased out of free allowances from 2013 to 2020 — but the threshold for exemption will be raised from 10,000 to 25,000 tonnes of annual carbon dioxide emissions.

In response to complaints from some countries, the committee proposed that some industries will still be eligible for 100% free allowances. But the exempt sectors will only be identified after the international climate summit in Copenhagen in 2009.

J. L. STANFIELD/NATIONAL GEOGRAPHIC/GETTY IMAGES

Europe's industry is facing tough emissions rules.

Cassini makes successful swoop on Enceladus



NASA/JPL/SPACE SCI. INST.

Saturn's moon Enceladus hoves into view in this image taken just 47 kilometres above its surface.

On 9 October, the Cassini spacecraft passed within about 25 kilometres of the moon on its closest approach — the nearest it has come to any moon during the mission. The trajectory took it straight through the icy plume that shoots out, geyser-like, from near the south pole of Enceladus.

This time, unlike during a close approach in March, the cosmic-dust analyser aboard Cassini worked properly and collected data on the particles of ice and water vapour that make up the geyser.

Harvard gift paves way for bioengineering institute

Harvard University announced last week that it has received a $125-million donation from business school alumnus Hansjörg Wyss to found a bioengineering institute. The donation is the largest from a single donor in the university's 372-year history.

The Wyss Institute for Biologically Inspired Engineering, to be headed by Harvard Medical School cell biologist Donald Ingber, will be located in the university's new $1-billion science complex in the Allston neighbourhood of Boston, Massachusetts. The complex is still under construction, but will eventually also host the Harvard Stem Cell Institute, in addition to several other programmes.

The new bioengineering institute will tackle interdisciplinary projects such as repairing diseased tissues and creating self-organizing materials and devices, and will be the home of the university's synthetic-biology programme.

'Manipulated' stem-cell paper faces retraction

The University of Minnesota has asked the journal *Blood* to retract a high-profile paper on adult stem cells following a university investigation. The paper reported that mesenchymal stem cells isolated from adult bone marrow could generate a surprising number of tissues (M. Reyes *et al. Blood* 98, 2615–2625; 2001), but other labs had trouble replicating that work. The investigation concluded that the paper included manipulated images.

Lead author Catherine Verfaillie, now director of the Stem Cell Institute at the Catholic University of Leuven in Belgium, was cleared of academic misconduct but blamed for insufficient oversight. That suggests the blame rests with the only other scientist under investigation, Verfaillie's graduate student Morayma Reyes, now at the University of Washington. However, the findings regarding Reyes cannot be released because of privacy laws. Reyes says that she made "honest unintentional errors".

Problematic images were also identified in a patent and in articles published by the *Journal of Clinical Investigation* and *Nature*. But the university did not find sufficient evidence of misconduct in these incidents.

The University of Washington says it may decide to investigate Reyes.

UN agency warns of biofuel production implications

The production of biofuels, which has tripled between 2002 and 2007, has significant economic, social and environmental costs. So says *The State of Food and Agriculture 2008* report from the Food and Agriculture Organization of the United Nations (FAO).

Biofuels make up just over 1% of the world's transport fuel consumption, but "this figure is expected to double over the next decade", says Terri Raney, senior economist at the FAO and co-editor of the report.

The report recommends that small farmers are given help to increase the types of market at which they can sell their crops. It also backs the further development of second-generation biofuels, from woods and grasses, which are not yet commercially available but could result in fewer greenhouse-gas emissions.

Is China ready for GM rice?

In an effort to avoid a food crisis as the population grows, China is putting its weight behind genetically modified strains of the country's staple food crop. **Jane Qiu** explores the reasons for the unprecedented push.

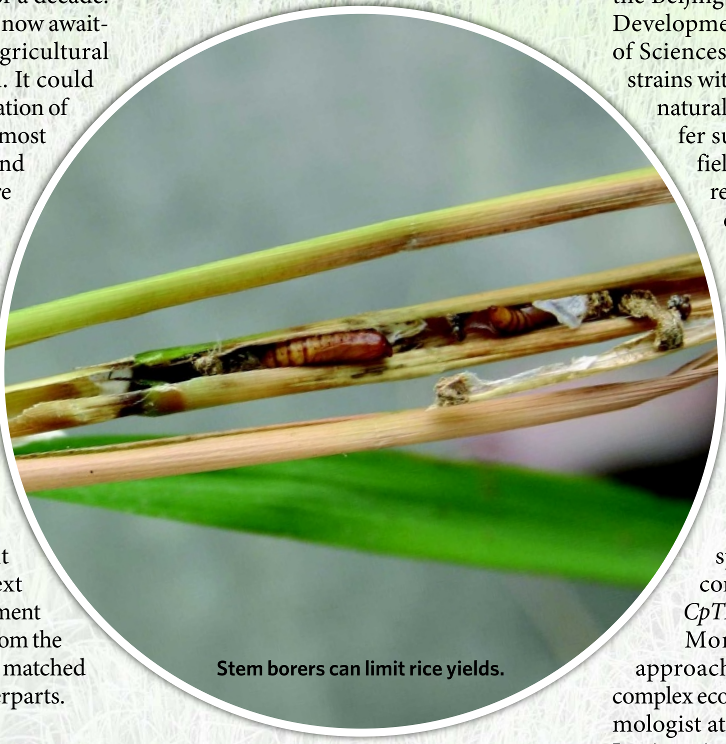
WANG FENG

In a paddy field 30 kilometres south of Fuzhou, the capital of China's Fujian province, Wang Feng is surveying a massive green and yellow chessboard before him. Wang, a rice researcher at the Fujian Academy of Agricultural Sciences, and his colleagues have been developing genetically modified (GM) rice strains to resist pest infestation, and have been testing in these plots for a decade. Two strains from Wang's team are now awaiting regulatory approval by the agricultural ministry for commercial growth. It could represent the largest commercialization of a GM foodcrop. Rice is a staple for most of the country's 1.3 billion people and a primary source of calories for more than half the world's population.

China's population is set to top 1.45 billion by 2020, and the country needs to increase grain production by about 25%, a daunting task in the face of increasing urbanization, industrialization, farmland reduction and the efflux of rural workers to the cities. The Chinese government has latched on to transgenic plants as a solution, rolling out a major research and development initiative on GM crops for the next 12 years, including a sizeable investment of 25 billion yuan (US$3.7 billion) from the central government and additional matched funding from its provincial counterparts.

The bigger picture

Like GM initiatives elsewhere, such as in the United States, the move has drawn its share of criticism, with concerns being raised about the practicality and safety of such a push. Scientists warn that a single-minded focus on genetic engineering to enhance pest resistance misses the bigger picture of how to address agricultural production. China is the world's largest rice producer, weighing in with nearly 200 million tonnes, and several observers fear that introducing GM rice could endanger the food supply and the environment. "The consequences would be unthinkable if large-scale cultivation of GM rice were not properly regulated," says Xue Dayuan, chief scientist on biodiversity at the Nanjing Institute of Environmental Sciences. But in a country where policies are rarely a matter of open debate, government officials warn that the scale of the impending food shortage makes further delays an unaffordable luxury. "This is the only way to meet the growing food demand in China," says Huang Dafang, former director of the Biotechnology Research Institute of the Chinese Academy of Agricultural Sciences (CAAS) in Beijing.



Stem borers can limit rice yields.

Wang is optimistic that his group's pest-resistant GM rice will help lead the way. In April, the team planted alternating squares of conventional rice crops and crops genetically modified to produce an insecticidal toxin made by the bacterium *Bacillus thuringiensis* (*Bt*) gene and the cowpea trypsin inhibitor (*CpTI*) gene. In the absence of chemical pesticides, *Bt*/*CpTI* rice thrived, whereas the conventional plants withered, resulting in the chessboard pattern of alternating colours. Wang pulls the top from one of the non-transgenic plants. Unrolling its leaves and splitting its stem, he reveals the insecticides' primary target, the stem borer.

WANG FENG

Wang says stem borers can affect 3.3 million hectares of rice fields, resulting in a 5% loss in yields at a cost of 10 billion yuan a year. According to Zhu Zhen, a geneticist at the Beijing-based Institute of Genetics and Developmental Biology, Chinese Academy of Sciences (CAS), who developed the rice strains with *Bt* and *CpTI* genes, there are no naturally occurring strains that can confer such resistance. After ten years of field testing at a dozen locations, the researchers are confident that farmers would use less pesticide with GM rice strains[1].

Plagued by pests

But David Andow, an entomologist at the University of Minnesota in St Paul, says he is unconvinced. In the past few decades, the stem borer has been overtaken by another pest, the brown planthopper (*Nilaparvata lugens*), which wreaks havoc every spring and has become the main concern of farmers in Asia. *Bt* and *CpTI* toxins have no effect on the insect.

Moreover, many simply see GM approaches as ham-fisted in the face of complex ecologies. Kong Luen Heong, an entomologist at the International Rice Research Institute in Los Baños, the Philippines, calls pest-resistant GM crops a short-term fix for long-term problems caused by crop monoculture and overuse of broad-spectrum pesticides. "Pests thrive where biodiversity is at peril," says Heong. "Instead of genetic engineering, why don't we engineer the ecology by increasing biodiversity?"

Indeed, such ecological engineering has proved beneficial. Zhu Youyong, president of the Yunnan Agricultural University in Kunming, and his colleagues have found that growing a mixture of rice varieties across thousands of farms in China could greatly limit the development of rice blast — a fungal rice

disease — and boost the yield[2]. They have also tested similar practices using different crops and found beneficial effects.

Although GM crops are, in principle, compatible with such an ecological approach, it requires management that has proved hard to achieve within China's agricultural system, which is based on small-scale cultivation, says Xue. Although *Bt* cotton, the biggest GM crop produced in China to date, has put the cotton bollworm (*Helicoverpa armigera*) under control, the population of secondary pests, such as mirids, has risen since 2001. That has led to increased pesticide use[3,4], but still at levels lower than pre-1997, when the cotton was introduced.

Safety concerns

Debates have also flared because of rice's central role in the Chinese diet. One concern has been that antibiotic-resistance marker genes used in the derivation of the transgenic plants could invariably be taken up by naturally occurring gut bacteria and lead to resistant, pathogenic strains. Both of Zhu Zhen's *Bt/CpTI* rice lines and other *Bt* strains developed elsewhere are free of such markers.

The GM plants must also be shown to be non-allergenic. Composition tests and studies assessing toxicity in non-human animals allow the developers to claim that the GM rice varieties are "substantially equivalent" to unmodified counterparts apart from the target-gene expression. But for food eaten three times a day by a billion people, short-term animal studies aren't enough to measure equivalence. "If there were a health risk, we would be heading for a major disaster," warns Liu Bing, an expert on science and society in Tsinghua University in Beijing.

Another concern is the potential environmental consequences of transgenes escaping from GM rice to its unmodified crop counterparts through cross-pollination. Several escapes have occurred around the world, including releases of unapproved GM crops such as rice and corn into human consumption streams. For example, in 2006, the European Union halted imports of US rice when an unapproved strain was found in the food supply. Trade resumed, but the problem of accidental cross-pollination, which is thought to have caused the contamination, is one that has not yet been solved. The consensus seems to be that perfect prevention of such events is impossible[5].

> **"Using GM rice is the only way to meet the growing food demand."**
> — Huang Dafang

Lu Baorong, a biodiversity researcher at Fudan University in Shanghai, is concerned about gene flow from GM rice to its wild or weedy relatives. Wild-rice plants are undomesticated strains, whereas weedy rice, which is characterized by its seed scattering and dormancy, is thought to originate from rice crops as a result of mutations. Lu's team and another group have shown that the rate of gene flow from GM strains to wild and weedy rice is 3–18% and 0.01–0.5%, respectively[6,7].

"What is most worrying is that such gene flow is cumulative," says Lu. Although rice crops are harvested at the end of the season, wild and weedy rice carrying transgenes would continue to reproduce, allowing the genes to spread, subject to selection.

This could threaten the biodiversity of wild rice, which provides a valuable gene pool for rice breeders (see 'The panda of the plant world') but is already at the brink of extinction in China. In addition, weedy rice with pest-resistant or other fitness genes could have a greater capacity to infest rice fields, causing yield loss. However, Lu says that these are not inevitable consequences of large-scale cultivation of GM rice. "Proper regulation is the key," he says.

A regulatory mess

But regulation, says Xue, is where the majority of problems lie. "Field testing is one thing, but the reality is quite another." Although China has had biosafety regulations for GM crops since 1996, their implementation has proved uneven — a fact that most people approached by *Nature* acknowledged. In some provinces, such as Xinjiang, farmers began large-scale cultivation of *Bt* cotton long before approval was granted, says Xue. In several cases, *Bt* cotton strains were grown without proper labelling, some of which were experimental strains from research institutes.

Cross pollination and labelling slip-ups could be disastrous for China's exports of rice and rice-related products. And proper regulation of GM crops is crucial for delaying the emergence of resistant pests. Many crops, such as cotton and rice, are grown as a monoculture in China, which would select pests that are resistant to the toxins. One way to avoid this from happening is to use seeds that produce high toxin levels; another way is to set aside some land near GM-crop fields for its unmodified counterparts, which would serve as a 'refuge' for insects.

This 'high-dose and refuge' strategy, which

J. QIU

The panda of the plant world

In the late 1960s, a seemingly ordinary wild plant attracted the attention of a young researcher at an agriculture college in Qianyang, Hunan province. It was a primitive relative of rice cultivars that did not produce pollen. The young man was ecstatic as this was the species he had long searched for and would allow him to generate hybrid rice with higher yields — thanks to heterosis, or hybrid vigour, in which the first generation of a cross between two crop varieties grows better and produces more grains.

The researcher, Yuan Long-ping, now considered the father of hybrid rice in China, has since led efforts producing hundreds of hybrid varieties, some of which increase yields by as much as 20% compared with inbred varieties. As wild rice species are highly adaptable, researchers continue to benefit from their diverse gene pools and have identified desirable traits such as resistance to drought, diseases and pest infestation.

"Wild rice species are extremely valuable for improving rice cultivation," says Xue Dayuan, chief scientist on biodiversity at the Nanjing Institute of Environmental Sciences. But increasing urbanization and industrialization has reduced natural habitats for wild-rice species to a tenth of the area they covered 30 years ago — despite the fact that the Chinese government classified wild-rice species as category II endangered species in 1992 — says Lu Yonggen, former president of the South China Agriculture University in Guangzhou, Guangdong province.

At the moment, only two conservation gardens, in Guangdong and Guangxi provinces, are dedicated to wild-rice protection, and there, species from other parts of the country are cultivated. But the controlled 'static' conditions of cultivation results in slower evolutionary rates, says Lu. It is crucial, therefore, to establish natural protection zones in the handful of provinces that still have the species. "Only then would they continue to thrive in the wild and become stronger and more genetically diverse plants," he says. Fondly known in China as the 'panda of the plant world', wild rice species "would slide into extinction if no immediate actions were taken", warns Lu. **J.Q**


Chinese researchers are working to reduce pest infestation in the paddy fields of Fujian.

J. QIU

has been widely adopted in countries that grow GM crops, is difficult to implement in China. Many Chinese farmers exchange seeds with each other or buy cheap seeds from illegal dealers, and end up growing cotton plants with low levels of *Bt* toxin. In addition, as agricultural practices in China are based on small-scale cultivation by individual families with limited resources, *Bt* cotton plants are grown without refuge areas.

Fortunately, the cotton bollworm also attacks crops such as wheat, corn, soya beans, peanuts and vegetables, which are grown next to the *Bt*-cotton fields and offer a safety valve against resistance[8,9]. "This is unlikely to happen with *Bt/CpTI* rice because the stem borer feeds only on rice," says Heong. "Therefore, setting aside refuge areas is absolutely essential."

Behind closed doors

Worryingly, many of the stakeholders, including farmers, bioethicists and environmental groups, aren't being involved in the biosafety evaluation process as spearheaded by the agricultural ministry. The country hails GM rice as a magic bullet for food-production problems and few dissenting opinions are heard. "The whole process is rather opaque," says David Just, an economist at Cornell University in Ithaca, New York. "China is trying very hard to keep the lid on." Experts who express their concerns are often sidelined. Xue, for example, has been repeatedly excluded from biosafety committees that are assessing GM crops.

Despite these concerns, China does need to find a way to feed its swelling population. Although it has instigated plans to prevent further reduction in farmlands, boosting grain production remains the key to food security. Still, the single-minded focus on genetic modification seems misguided to many.

"Genetic-modification technologies just treat the symptoms rather than dealing with the causes," says Hans Herren, president of the Millennium Institute in Arlington, Virginia, and a co-chair of the International Assessment of Agricultural Knowledge, Science and Technology for Development (IAASTD). According to a report released by the IAASTD in April, the main challenges faced by agricultural development around the world are soil fertility, water management and climate change[10]. "Life in the soil is gone after decades of heavy use of pesticides, herbicides and chemical fertilizers," says Manuela Giovannetti, a soil microbiologist at the University of Pisa in Italy. Herren agrees: "Without a concerted global effort to restore soil fertility, genetic modification would be futile."

"The consequences would be unthinkable if large-scale cultivation of GM rice were not properly regulated."
— Xue Dayuan

Xue says he recognizes the potential of genetic modification, but is concerned that huge investment in the technologies — as with China's new initiative — would sap already dwindling attention from improving traditional plant-breeding technologies and conventional farming practices. However, GM strategies have a strong draw for keeping China competitive at the cutting edge of agriculture. A report by the International Service for the Acquisition of Agri-biotech Applications estimates that biotech rice could deliver benefits of $4 billion per year for China[11].

"What is behind all this might be about who controls germ plasm and who owns intellectual-property rights," says Andow. The scale of the effect that commercial GM rice could have on China and the rest of the world argues for caution. Nevertheless, many interests within the country say that the time to act is now. Huang puts it bluntly: "We cannot afford to think too far ahead but must tackle the present issues." ■

Jane Qiu writes for Nature from Beijing.

1. Huang, J., Hu, R., Rozelle, S. & Pray, C. *Science* **308,** 688–690 (2005).
2. Zhu, Y. *et al. Nature* **406,** 718–722 (2000).
3. Wang, S., Just, D. & Pinstrup-Andersen, P. *Int. J. Biotechnol.* **10,** 113–120 (2008).
4. Wang, Z. *et al. Agric. Sci. China* (in the press).
5. Ledford, H. *Nature* **445,** 132–133 (2007).
6. Wang, F. *et al. Plant Biotechnol. J.* **4,** 667–676 (2006).
7. Shivraina, V. K. *et al. Crop Protection* **26,** 349–356 (2007).
8. Wu, K.-M., Lu, Y.-H., Feng, H.-Q., Jiang, Y.-Y. & Zhao, J.-Z. *Science* **321,** 1676–1678 (2008).
9. Qiu, J. *Nature* doi:10.1038/news.2008.1118 (2008).
10. Beintema, N. *et al. International Assessment of Agricultural Knowledge, Science and Technology for Development: Global Summary for Decision Makers* (IAASTD, 2008); available at http://www.agassessment.org/index.cfm?Page=IAASTD%20Reports&ItemID=2713
11. James, C. *Global Status of Commercialized Biotech/GM Crops: 2007* ISAAA Brief No. 37. (ISAAA, 2007); available at http://www.isaaa.org/Resources/Publications/briefs/37/executivesummary/default.html


WANG FENG

NATURE|Vol 455|16 October 2008

ILLUSTRATION: P. DISLEY

Sam Ting's last fling

The International Space Station's one chance of scientific greatness rests on a high-profile refugee from the world of the particle accelerator — but is it too long a shot to be worth taking? **Eric Hand** reports.

In 1994 Dan Goldin, then the administrator of NASA, was on the look out for sexy scientific ideas. In particular, he needed something that would endear the International Space Station — a merger of America's earlier space station plans with those of Russia — to scientific sceptics unimpressed by the experimental opportunities it offered. Roald Sagdeev, the former director of a major Soviet space research institute, mentioned to Goldin that Samuel Ting, a Nobel prizewinning particle physicist, was toying with ideas for a spaceborne magnet that would sift antimatter from the stream of particles from beyond the stars. "He said, 'Okay, where is this guy? I want to see him immediately'," recalls Sagdeev.

By 1995 Ting, a professor at the Massachusetts Institute of Technology in Cambridge, had won Goldin's agreement that NASA would give his Alpha Magnetic Spectrometer (AMS) space on the station, and a shuttle flight to get it there. Part of that agreement was that the agency would not have to pay for the AMS to be built. Instead Ting got the Department of Energy (DOE), which funds US particle physics, to oversee the AMS — and undertook to round up most of the funding from foreign governments.

That he did. Thirteen years on, and by some accounts US$1.5 billion down the line, a vast team of physicists from 16 countries has put all but the finishing touches to the AMS. But since returning to flight after the loss of the Columbia in 2003, the shuttles — due to retire in 2010 — have been devoted almost entirely to completing the space station. A dedicated AMS flight was dropped from the manifest in 2005, and reinstating one seemed, until recently, out of the question.

Yet Ting, 72, has soldiered on. A Nobel prize and decades running huge particle-physics collaborations give him considerable heft; few physicists have the clout to get an Italian foreign minister to plead their case to US secretary of state Condoleezza Rice. And the presidential campaign has provided extra leverage to the weight Ting brings to bear on lifting the AMS into the sky. An extra shuttle flight is a nice thing to promise Florida voters worried about jobs that will disappear when the shuttle is grounded for good, and both candidates for the presidency have recently made such promises. Congress has sent on to the White House a bill authorizing Ting's shuttle flight.

Over tea in an empty but opulent dining room at the Mayflower Hotel — Ting's campaign headquarters in Washington DC — he folds himself into a chair and cracks open a laptop full of documents, charts, even old clips from *The New York Times* lauding his past achievements. "I know all the technical details," he says. "I'm the one responsible if something goes wrong. I don't do anything else but this." Every day for 13 years, it has been his focus, an all-consuming passion and worry; every day except, that is, for ten days in 1998, when the prototype AMS-01 was flown on the shuttle Discovery. With the experiment floating weightlessly, gathering data from particles

passing through its doughnut-shaped magnet, Ting felt himself relax. "For the first time, I realized that if something went wrong, there was nothing I could do."

Ting's case, bolstered by slide after slide from the laptop, is that the AMS will open up a new spectrum to astronomy: that of charged particles. Antimatter left over from the Big Bang is an imagination-grabbing example of the sort of thing it might find. But only an example: the real need for the AMS, in Ting's mind, is to discover the utterly unforeseen. One of his slides lists the originally cited aims of a huge range of 'big' science projects over the past 50 years — and the discoveries for which they are famous. The two are always different. "What you will see," he says, in his slow, soft voice, "it's hard to predict."

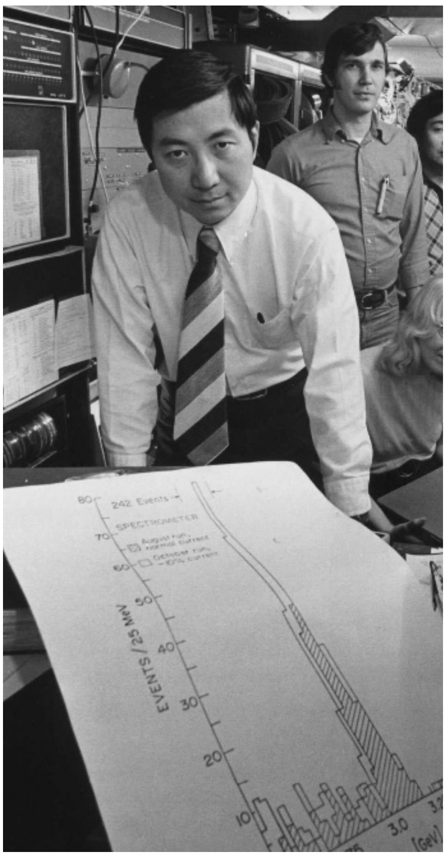
"For the first time, I realized that if something went wrong, there was nothing I could do."
— Samuel Ting

The numerous opponents of the AMS, however, think that they can predict the project's results — and that they are likely to be relatively underwhelming. No current theory leads them to expect the presence of antimatter nuclei in space of the sort Ting talked to Goldin about. There are other things, they say, on which the money that would be needed to launch the AMS could be better spent — things the astrophysical community has evaluated and prioritized. But Ting will have none of it. In 1976, his Nobel lecture offered a tale of careful experiment proving theorists wrong. Experiments are most meaningful when they disagree with theory, he says, with an emphatic tone that brooks no dissent. "The advancement of physics depends on you destroying other people's perceptions."

Sam I am

Ting's record backs up his belief in the transformative value of daring experiment. "By being just that much more clever or careful than everybody else, Sam's able to get stuff out that other people missed," says John Ellis, a theorist at CERN, Europe's particle-physics facility near Geneva. It was through such painstaking measurement that Ting won his Nobel prize in the first place. At Brookhaven National Laboratory in Upton, New York, Ting managed to pick the signature of a new particle out of a very messy energy spectrum with almost-over-the-top levels of instrumentation, and a monumental insistence on thoroughness. When Ting shared the 1976 Nobel Prize in Physics for the discovery, the prize committee described his feat as being "like hearing a cricket near a jumbo jet".

As Ting's reputation grew, so too did the size and scope of his experiments. By 1983, Ting was leading what became the largest physics collaboration in the world: the L3 experiment at CERN's Large Electron Positron (LEP) collider, the first machine to occupy the vast tunnel that now houses the Large Hadron Collider (LHC). With almost 500 physicists, L3 employed more people than any of the other three experiments spaced around the LEP ring. It was also remarkably international, cementing in place Ting's reputation as one of his discipline's great fixers. An American born in Michigan and raised in Taiwan, he got China to supply special crystals to the detector — its first foray into high-energy physics at that level — while getting the Soviet Union to contribute an Eiffel Tower's worth of iron. His leadership approach was not exactly democratic, but he accomplished things quickly and decisively, says David Stickland, a CERN physicist who worked on L3. He recalls a time when Ting was asking the DOE for an upgrade to the LEP experiment. The DOE rejected his proposal. "Sam just stood up and said, 'I reject your rejection,'" recalls Stickland. "He really operates outside the norms that way."



Sam Ting at Brookhaven displaying the results that won him a Nobel prize in 1976.

BROOKHAVEN NAT'L LAB.

L3 ran on the LEP until the accelerator was closed down in 2000. But Ting's attempts to build something even bigger were thwarted. The coalition he put together for a detector to grace America's planned Superconducting Super Collider involved 1,000 scientists from 90 institutions in 13 countries; but it suffered internal strains, and the collider's management rejected it in 1991. Pivoting back to CERN, Ting suggested putting a revamped L3 onto the LHC. Unusually for him, this was a comparatively cheap proposal. But CERN rejected it.

Licking their wounds, Ting and his close collaborators began to think about something smaller and simpler — a break from the demands of giant collaborations, according to Ulrich Becker of the Massachusetts Institute of Technology, who has worked with Ting since 1965. The idea for the AMS was born on a coffee break from L3 work. "I had this dream to build an experiment that would have fewer than 100 collaborators and could fit on a table," Becker says. The idea endured: the scale didn't. The project has involved not 100 but 500 scientists, from 56 institutions. At 7 tonnes, it would need the sort of table a minibus can be parked on. Its extraordinary 0.86-tesla magnetic field is 17,000 times bigger than Earth's and five times greater than a sunspot's. If the AMS didn't spend half its energy cancelling out the field lines that would otherwise stray beyond its confines, the space station wouldn't stay stationary very long. "Sam Ting doesn't like to do small things," says Becker.

Big-Bang refugees

The AMS team sees celestial charged particles, also called cosmic rays, as a way to look into a problem that particle physics has not yet solved on its own terms: why is the Universe mostly matter, not antimatter? Processes that favour matter over antimatter clearly played a role in the Big Bang. How those processes played out, though, is still something of a mystery. The contribution the AMS team hopes to make to this debate is to see whether the Universe's bias against antimatter is as complete as is normally assumed.

If any antimatter did escape the annihilations of the early Universe, then there could be stray antiatoms still around today. And if an antihelium nucleus — the lightest antiatom that can't have been formed by any known process since the Big Bang — were to pass through the central void in the AMS's doughnut, its mass and charge would be immediately revealed by the way its trajectory bent in that awesome magnetic field. The distinctive curve of antihelium would be a revolutionary discovery. That of any heavier antimatter would be simply mind-boggling — a sign of antigalaxies

and antistars somewhere far off and as yet unobserved, their nuclear fires fusing together the antimatter equivalent of the stuff of which Earth, and humans, are made.

Antimatter does not have to be primordial. Its lightest particles can be made in various ways, some known and comparatively prosaic, some fanciful and as yet unseen. Some theories hold, for example, that the decay of small black holes might produce antineutrons stuck to antiprotons. Other processes can make these too, but those coming off black holes would move peculiarly slowly — something the AMS's sensitivity to mass, charge and speed would pick up. Cold dark matter (CDM) — hypothetical particles thought to account for most of the mass of the Universe — could be a contemporary source of distinctive antimatter too, in the form of high-energy positrons given off when the CDM particles decay. And then there's strange matter — matter made up from more types of quark than just the two basic ones that make up neutrons, protons and their antiparticles. The AMS could conceivably detect light, long-lived particles of this quark matter.

These wonders sold the idea to Goldin in the 1990s. But they have not convinced the astrophysicists who account for most of NASA's astronomical constituency. Once every decade the US National Academies produce a major report stacking up astronomers' research projects against each other. NASA uses these reports as prioritized shopping lists. The decadal report that came out in 2001 makes only a desultory mention of the AMS, treating it as if it was something outside the report's purview, destined to happen regardless. "The AMS was something that came out of high-energy physics

CERN

A technician checks the spectrometer.

CERN

Sam Ting (right) surveys the AMS experiment with representatives from China.

as a big project at a high level," says Thomas Gaisser, of the University of Delaware in Newark and the panel chair responsible for reviewing projects such as the AMS for the decadal report. "The people who had been working in the cosmic-ray fields for a long time didn't like the competition, basically." The decadal report did not endorse the mission.

The antihero

In a letter from February of this year, Craig Hogan, chair of an astrophysics advisory committee to NASA, noted that the AMS also went unmentioned in a 2003 National Academies report that was specifically supposed to consider cross-community projects dealing with the nature of cosmic matter. "The overall health of the astrophysics programme is put at risk by any mission whose science value has not been transparently compared with other missions," wrote Hogan, director of the Center for Particle Astrophysics, at the Fermi National Accelerator Laboratory in Batavia, Illinois. Ting responds by pointing out that the AMS is not a NASA astrophysics mission. No NASA science money has been used, and no NASA scientists have worked on it. In terms of being reviewed for scientific merit he notes DOE reviews performed in 1995, 1999 and 2006, in addition to multiple reviews by European agencies.

> "Sam just stood up and said, 'I reject your rejection'."
> — David Stickland

Another line of attack is now opening up. When the AMS was first proposed, tests for antimatter in cosmic rays had hardly been tried: the new window on the Universe they offered was wide open. Now there are some data. BESS (Balloon-borne Experiment with a Superconducting Spectrometer) has looked for antimatter in cosmic rays on three high-altitude balloon flights around Antarctica; the PAMELA (Payload for Antimatter Matter Exploration and Light-nuclei Astrophysics) satellite was launched in 2006. Both, oddly enough, are cheap descendents of an earlier cosmic-ray experiment proposed for the US space station in the 1980s. Their verdict on primordial antimatter? It probably isn't there. "Will AMS provide any fundamental answers that BESS and PAMELA haven't?" asks John Mitchell, US principal investigator for BESS at NASA's Goddard Space Flight Center in Greenbelt, Maryland. "The answer is, probably, no."

This doesn't mean that the AMS has no chance of finding primordial antimatter: but the window has closed down a lot. And Ting's nippy little rivals are also making inroads into other science that the AMS could have had to itself. There are enticing, as yet unpublished, hints that PAMELA is seeing some of the high-energy positrons that might be expected from decaying cold dark matter (see *Nature* **454,** 808–809; 2008). If this turns out to be the case, the news for the AMS might be bittersweet. Its far greater power and spectral range would be excellent for further analysing those positrons, and so the case for launch would be strengthened. But rather than opening a new window on the Universe, its primary purpose would be reduced to that of follow-up to someone else's discovery.

In the meantime, as PAMELA gathers more data, the AMS just sits at CERN, where it was assembled. Over the summer, technicians calibrated it, blasting it with high-energy particles. At the end of this year, the AMS will be taken to a European Space Agency facility in the Netherlands, where it will be tested in

space-like conditions. As early as June 2009, the AMS could be ready to cross the Atlantic, tied down in a 747 Lufthansa cargo plane, to Kennedy Space Center on Merritt Island, Florida — if a shuttle were waiting.

Ting is still rueful, worried about the damage to US credibility in a project where foreign partners have footed most of the bill. "The US government made a commitment to fly it," he says. "This should have been thought about a long time ago — not after more than $1.5 billion has been spent." To some, this seems to be protesting too much. The material costs of AMS-01 were just $33 million. AMS-01 used a lower-field permanent magnet built by the Chinese for $600,000, far cheaper than the superconducting Swiss toroid in the grown-up version, and there have been many other improvements. But can they really account for a hike in price to $1.5 billion?

Simon Swordy, a cosmic-ray physicist at the University of Chicago in Illinois, takes what might be called a worldly view of the inflation: when a scientist initially sells a project, it should sound cheap; once it's built and ready to go, it's better to be expensive. "Ting wants to say, 'You've already spent $1.5 billion on this, you've just got to fly it'." Ting responds by saying that the $1.5-billion figure is an extrapolation of a $1.2-billion NASA estimate made in 2005. That put the material costs of the AMS at $179 million; the other billion was in overheads, facility costs, or salaries, reflecting in part NASA's shift to 'full cost accounting'. Ting also points out that this cost estimate was made before the AMS was bumped off its shuttle flight, which is hard to square with the suggestion that he needed an artificially hefty price tag to get the AMS into space.

Whatever the costs, though, they are now, as economists put it, sunk. They cannot be recouped whatever happens. In cost–benefit terms, the costs that matter are those needed to get to the end of the project, not those incurred since its beginning. An extra shuttle flight squeezed into a crowded 2010 would cost between $300 million and $400 million, according to a 2008 NASA estimate. Doing it later is an order of magnitude costlier: renewing shuttle contracts for the 2011 fiscal year would cost $3 billion–4 billion. Engineers have explored other options for getting the AMS up — reconfiguring it as a free-flying satellite, for instance, or using the European Space Agency's Automated Transfer Vehicle to get it to the station. But at this stage, those options are not feasible, says Mark Sistilli, NASA's programme manager for the AMS.

"By being that much more clever, Sam's able to get stuff out that others missed."
— John Ellis

Waste of space shuttle?

Shuttle managers need 18 months lead time to prepare a shuttle. If the schedule to retire the shuttles by October 2010 is kept, NASA managers say they need to know in early 2009. "Time is of the essence," says Sistilli. Hence Ting's regular visits to the Mayflower Hotel — part of a campaign that is meeting with new success. On 27 September, both houses of Congress passed a NASA authorization bill that specifically directs the agency to add another shuttle flight and to use it to take the AMS up to the space station. That bill is currently waiting for the president's signature or veto. Although the White House has so far been against such a flight, a veto is not seen as likely.

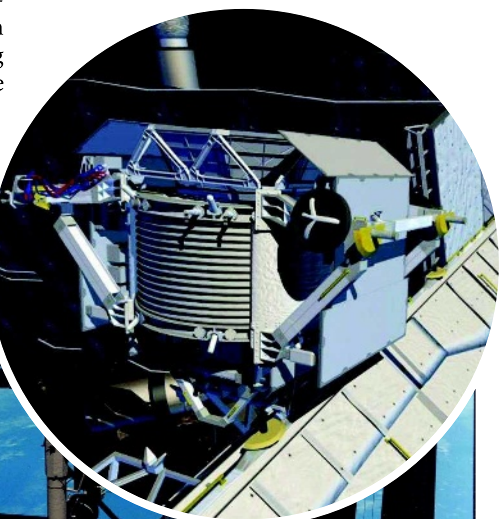
The worry for NASA astrophysicists is that if the bill authorizing the extra flight is enacted into law, it is by no means certain that Congress's appropriations committees will come up with extra money to pay for it. If that were the case, the NASA astrophysics budget may have to pay for the AMS's ride, says Jon Morse, NASA's astrophysics division director. "That is the risk that currently exists." If you want a bottom-line explanation of why the astrophysicists have not welcomed what they see as an un-peer-reviewed interloper, look no further. Congress could make the AMS flight a zero-sum game in which the $300 million–400 million shuttle flight comes out of the $1.3 billion astrophysics budget, and other missions will be cancelled or delayed.

The AMS team sees it from a different perspective. Ting has transcended the zero-sum game, using his political muscle to round up money, piecemeal, from the many international partners that would not otherwise have been available. The AMS scientists see the extra shuttle flight as not only the fulfilment of an original obligation, but also small compared with the $1.5-billion overall price tag. With regards to the potential robbing of the astrophysics budget, Ting says the original agreement was not with NASA's science division. "There's no reason we should take money from them." That said, he is not in a position to make that call.

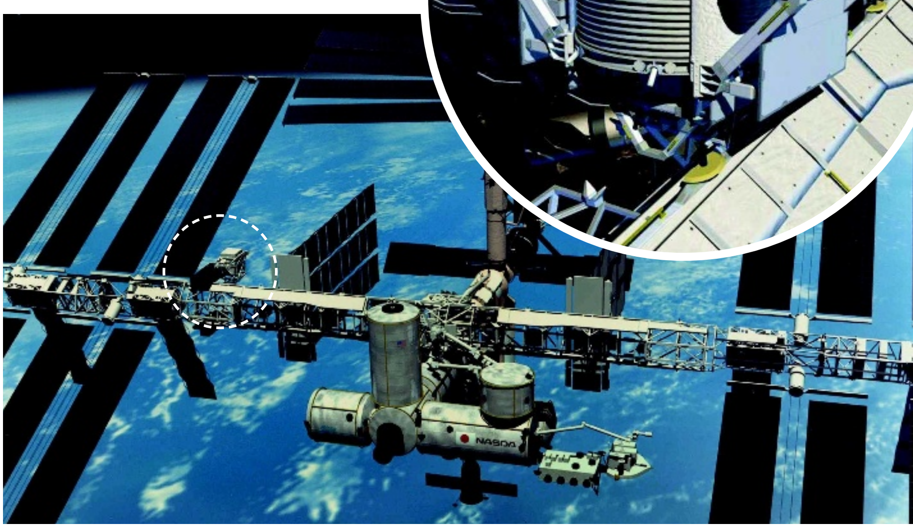
And just as $300 million is small, say Ting and his colleagues, when compared with $1.5 billion, so $1.5 billion is small compared with $100 billion — a frequently bandied about figure for the total costs of the space-station programme. So far, they say, that astronomical sum has bought very little in terms of science: for 1% of the total, the AMS might go some way to redeeming that. "Can you name me one single important discovery done there?" asks Roberto Battiston, deputy principal investigator for the AMS. He recalls reading a list of 12 International Space Station highlights in 2007 that NASA published. It included the clubbing of a golf ball during a space walk, a stunt sponsored by a Canadian company; it did not include any science. "Honestly, I think the station is there for something more than that." ■

Eric Hand covers physical sciences for *Nature* from the Washington DC office.

AMS-02 ROMA



NASA/CERN



A new home? This mock-up shows the AMS *in situ* on the space station — if it manages to get there.

The Standing Rock Sioux Reservation runs a field course for budding palaeontologists in the tribe.

G. GRELLET-TINNER

School of rock

Native Americans want to claim fossil resources found on their lands. **Rex Dalton** looks at how tribes and palaeontologists are working together to avoid bitter ownership disputes.

In the heart of America's dinosaur country, the relationship between Native Americans and outside palaeontologists has always been tense. In the 1890s, the battle between white settlers and Native Americans was barely over when legendary fossil-hunter Edward Cope arrived to prospect for bones in the grassy hills that make up the Standing Rock Sioux Reservation in North and South Dakota. A century later an amateur palaeontologist digging in the neighbouring Cheyenne River reservation unearthed the largest and best-preserved *Tyrannosaurus rex* known, setting off an epic legal battle involving the research institute she worked for, the Sioux rancher who owned the land on which the fossil was found, and local tribes.

But that long-standing tension was nowhere to be seen this July. On the Thunder Hawk Ranch on the Standing Rock reservation, near the sacred grounds where Chief Sitting Bull was born and died, Native American students came together for what is believed to be the nation's first palaeontological field school conducted by a tribe. Led by Swiss-born palaeontologist Gerald Grellet-Tinner, the students worked just as Cope did a century ago, unearthing fossils under a punishing sun broken only by torrential rain.

If this field school is any indication, Native Americans could be on the verge of instituting sweeping changes that will resonate throughout the palaeontological community. Some tribal members, such as those who ran the dig site this summer, want to focus on the history and fossil resources of their land. Others, though, are seeking ways to protect their palaeontological heritage more aggressively.

Two decades ago, Native Americans pressed for new laws covering ancient human remains. The resulting legislation, the 1990 Native American Graves Protection and Repatriation Act (NAGPRA), transformed US archaeology. NAGPRA required that any remains be returned to the custody of tribes who could prove they had a 'cultural affiliation' with them. Thousands of items have since been repatriated, prompting archaeologists to complain that their museum collections are being depleted of specimens for scientific study.

Land rights

Now, some tribes are looking to extend that concept to fossils. Next week, a committee in the Nebraska legislature will hold a hearing about whether new legislation is needed to protect palaeontological resources on tribal lands. Nebraska's decision is important because the state contains large expanses of reservation land — and as such it often pioneers laws that involve Native Americans. For instance, it enacted a law protecting Native American human remains even before the federal NAGPRA was passed.

The Nebraska drive was spurred by Lawrence Bradley, a doctoral student in geography at the University of Nebraska in Lincoln, who was raised by the Lakota people. Bradley openly acknowledges that he would like to see

the repatriation of fossils taken by major East Coast museums a century ago[1].

Others are simply trying to improve the often-rocky relationships of today. In Standing Rock, for instance, the Dakota/Lakota people are now seeking the return of specimens taken from tribal lands by fossil hunters (see 'Fossil fray'). And in New Mexico, leaders of the Pueblo of Jemez community charge that a scientist from a state museum tricked them into getting access to the reservation, but didn't train a Native American student as they expected in return.

In 2004, tribe member Kevin Madalena agreed to help Spencer Lucas, both at the New Mexico Museum of Natural History and Science in Albuquerque, get access to the Jemez reservation to collect geological samples. But Madalena says that Lucas gave him very little instruction, cutting him out of the possibility of co-authoring the resulting article[2]. Lucas, who has been criticized before for the way he deals with students[3], says that he is surprised to hear of the criticism.

But for the young Madalena, the real lost opportunity was for his tribe and its resources. "It is imperative to educate native communities and non-natives about the importance of guarding palaeontological resources," he says.

At Standing Rock, they are working on that bond. Last year, the governing tribal council approved a 24-page Paleontological Resource Code, which set out provisions for a palaeontology office, defines agency responsibilities, created permitting requirements, and affirmed tribal ownership rights of specimens. The tribe is thought to be the first in the United States and Canada to adopt such a code, authorities say.

"The tribe has fantastic palaeontological resources, eager students and the financial support of its leaders."
— Gerald Grellet-Tinner

The strict rules have taken a while to be taken up by the reservation, which encompasses nearly a million hectares on the west side of the Missouri River. For years, the tribe tried to curb looting by tightly controlling access to the reservation; tribal members were also told to stop picking up any fossils they might find.

But in 2004, Gale Bishop, then director of the geology museum at the South Dakota School of Mines and Technology in Rapid City, helped the US Bureau of Indian Affairs and the tribe to collaborate with his museum and the Indian Affairs bureau to assess and develop the palaeontological resources. After Bishop retired in 2006, however, the university and the tribe began to differ over how to complete the survey.

Then Grellet-Tinner arrived. The lanky outdoorsman came from Switzerland by way of Texas, Los Angeles and Brazil. Trained as a gemologist, he later turned to palaeontology and explored rich fossil beds in China and Argentina. By the time he got to South Dakota's School of Mines and Technology, his experience of working with different cultures helped him smooth over the troubles between the tribe and the university. Grellet-Tinner straightened out the paperwork and contracted with the tribe to help develop its palaeontology programme.

"The tribe has a tremendous opportunity here — fantastic palaeontological resources, eager students and the financial support of its leaders," he says. The Standing Rock council spent US$50,000 to run the field school this summer, including a second week-long session for half-a-dozen tourists.

Tribal council member Henry Harrison credits Grellet-Tinner with turning the palaeontology programme around. The Standing Rock tribe now runs the programme from its headquarters in Fort Yates, North Dakota. The 6,000-member tribe is also building a fossil-preparation laboratory at its Sitting Bull College and hopes to build a palaeontological museum to bring not only tourists and economic development, but also scientists to study specimens found on the reservation.

The land contains a wealth of material. The field site is on a working cattle ranch, which is located a 30-kilometre drive down gravel roads from the main highway. Rocks here expose both the Hell Creek Formation — one of the most famous sedimentary beds for dinosaur fossils, dated at just over 65 million years old — and the slightly older Fox Hills Formation, at 68 million years old. Both sites record how the great Cretaceous seaway that once divided North America was receding at the time. "What is now Montana was exposed before the Dakotas," says Grellet-Tinner. "When Montana was a delta, the Dakotas were under water. When Montana was a plain, the Dakotas were becoming a delta."

With each rainstorm and snow melt, fossils erode out by the hundreds. "I found sites that were literally bone beds, with so many fossils you couldn't walk on the ground without stepping on fragments," says Bishop, who is now at Georgia Southern University in Stateboro. "There are several world-class deposits on that reservation."

Bone bounty

This July, students recovered and casted about 40 substantial specimens, which are now at the lab at Sitting Bull College. They include a possible new crocodilian species, along with sea shells that act as markers of the climatic conditions as the great seaway was receding. "I think this material will lead to some quality publications," says Grellett-Tinner.

To the budding scientists, the chance of contributing to a scientific article was more than they dreamed of in their camping tents. "I never imagined it would be like this; it was awesome," says Vida Dogskin, a Dakota/Lakota from Sitting Bull College. A single parent with four children, Dogskin is an environmental-science major now following an interest in dinosaurs. "I

R. DALTON

Gerald Grellet-Tinner has turned around relations between academics and the Standing Rock tribe.

Fossil fray

A South Dakota tribe and a Minnesota college are clashing over the rights to some important palaeontological specimens.

Five years ago, government officials and tribal leaders in South Dakota discovered that excavations on private land had extended onto a tribe-owned ranch. In all, fossil hunters had collected several thousand specimens, including a suite of duck-billed dinosaurs, or *Edmontosaurus*, ranging from juveniles to adults. And now the Native Americans want the fossils back from Concordia College, the institution in Moorhead, Minnesota, that helped collect many of them.

"We want them to do the Christian thing and return the fossils," says Steven Emery, counsel for the Standing Rock Sioux Reservation.

But for years Concordia's president, Pamela Jolicoeur, has refused all requests from the tribe to return the bones. Only now, on enquiries from *Nature*, has an attorney for the college, Bruce Quick, suggested that the tribe might get the fossils back. "If they continue to want them returned," he says, "they can have them."

The dispute began in June 2003, when officials at the US Bureau of Indian Affairs found Ronald Nellermoe, a biologist at Concordia, and his students excavating in a bone bed overlooking the Grand River. The officials discovered that a tractor digging on private land owned by the Schmidt family had extended excavations onto the Thunder Hawk Ranch owned by the Standing Rock tribe. They ordered that all digging be stopped, and it hasn't resumed since.

Nellermoe's dig ran alongside another one conducted by Adventure Safaris of Santa Maria, California, which sells excavation opportunities as a 'mobile creation-science camp'. Nellermoe, who says he is not a creationist, is also a past board member of the Grand River Museum, a creationist facility built by the Schmidts in the nearby town of Lemmon, South Dakota.

Tribal members suspect that some fossils displayed in the museum came from the reservation. Nellermoe denies that this is the case. The tribe has no proof, and hence has focused its demands on bones at Concordia, he says.

The tribe has repeatedly asked the US attorney's office for South Dakota to investigate and prosecute offenders. The officials refused.

In a letter to tribal council chairman Ron His Horse Is Thunder in August 2006, Jolicoeur admits that a loader was used to dig on reservation land. And Nellermoe acknowledges that an undetermined number of fossils had been unearthed from tribal land. He says he knows which fossils are from which site and they could "probably" separate out the reservation specimens. But "we never felt we had to do that," he adds.

Nellermoe has since given some of the *Edmontosaurus* specimens to other scientists for isotopic analysis[5], but his group has otherwise published little beyond some geological background[6].

Today at the Grand River Museum, a wall mural shows dinosaurs romping alongside bison. Staff at the creationist museum say that the dinosaur fossils on display are there to illuminate "extinction", not "evolution". According to them, dinosaurs just didn't make it onto Noah's ark in time. And so they display a model of the ark right next to the dinosaur fossils. **R.D.**

R. DALTON

The Grand River Museum pitches fossils as examples of extinction rather than evolution.

jumped in to learn as much as I can," she says. Dogskin picked up the trade so quickly that she was hired as a technician apprentice to help store and inventory the summer's finds.

For Matthew Wood, a member of the Seminole tribe originally from Oklahoma, the field school was a chance to fulfil a lifelong desire to study dinosaurs. A junior information-science major at Sitting Bull College, he is now looking at a career in palaeontology or a related field. "I thought I might learn a little, but I learned much more than I expected," says Wood. "It was all hands-on — an incredible experience."

The summer's finds are already heading for analysis elsewhere. Grellet-Tinner plans to send ammonite fossil samples from the Fox Hills Formation to the University of Claude Bernard Lyon I in France. Researchers there are preparing the specimens to be analysed in a mass spectrometer to measure their ratio of oxygen-16 to oxygen-18, which can indicate the temperature and depth of the water in which they formed. Grellet-Tinner says the team will also be analysing dinosaur bones — in part to try to determine whether dinosaurs were warm- or cold-blooded, which is still under debate[4].

Will students who participated in the field-school research be included in the article? "Of course," he says. And one day this may be the model for a new era of paleontological research on tribal lands — in which Native Americans collaborate with outside experts and share in the scientific analysis and credit. If so, it will be a long way from the days of the wars over bones on the Great Plains. ■

Rex Dalton is a reporter for *Nature*, based in San Diego.

1. Dalton, R. *Nature* **449,** 952–953 (2007).
2. Lucas, S. G., Krainer, K., Colpitts, R. M. *New Mexico Mus. Nat. Hist. Sci. Bull.* **31,** 101–117 (2005).
3. Dalton, R. *Nature* **451,** 510 (2008).
4. Amiot, R. *et al. Earth Planet. Sci. Lett.* **246,** 41–54 (2006).
5. Thomas, K. J. S. & Carlson, S. J. *Palaeogeogr. Palaeoclim. Palaeoecol.* **206,** 257–287 (2004).
6. Colson, M. C., Colson, R. O. & Nellermoe, R. *Rocky Mountain Geol.* **39,** 93–111 (2004).

CORRESPONDENCE

Britain should not abandon curiosity-driven research

SIR — I am pleased that the recent Wakeham Review of physics acknowledges the strength of UK astronomy and space science (B. Wakeham *Nature* **455,** 592; 2008). I also welcome the clear statement that basic research in the United Kingdom should be funded at a level needed to maintain the country's international standing.

In the wake of last year's science-funding crisis in the United Kingdom, in which the body responsible for funding UK research in astronomy, the Science and Technology Facilities Council (STFC), was hit with a £80-million (US$136-million) budget shortfall, Bill Wakeham and his team carried out a thorough review of physics and astronomy. They make clear and positive recommendations that I am largely happy to support, and that I hope the new science minister and the UK government will soon implement so as to avoid the funding problems we experienced last year. (Reprofiling within the STFC has helped with the projects that were threatened, but there is still a £25-million hole in grants funding.)

The good news from the Wakeham Review contrasts with a call last month from the former chief science adviser to the government, David King, for the best scientific minds to be redirected towards solutions to twenty-first-century problems such as climate change. To achieve this he suggests that less time and money be spent on space exploration and particle physics.

It is extraordinary that he should say this, in a BBC television interview, in the week that the Large Hadron Collider was switched on (see http://tinyurl.com/5cayt8 and http://tinyurl.com/4cm42p). The drastic budgetary measures in the STFC took place while King was still advising the government, and the United Kingdom continues to make contributions to the European Space Agency at little more than half the level of France and Germany.

Tackling climate change is certainly of great importance, but it is not clear that it must be done at the expense of other sciences. This approach would reduce the country's overall science base when it actually needs to increase. Tackling climate change not only requires scientific solutions but also has enormous political and social dimensions. A scientifically literate society will help.

Astronomy and space exploration have long been important in interesting and educating the public of all ages in science. In particular, these fields stimulate bright young people to study science.

King seems to want to move researchers from 'curiosity-driven' science, such as space exploration and particle physics, into 'driven' science. This does not make sense and is not how scientists are motivated.

Andy Fabian Royal Astronomical Society, Burlington House, Piccadilly, London W1J 0BQ, UK
e-mail: acf@ast.cam.ac.uk

Editorial comment should accompany hot papers online

SIR — Online publication of high-profile papers ahead of print is a great facility for speedily alerting the community to important research results. But these advance online publications rarely carry an accompanying editorial comment, either in *Nature* or in other journals. In traditional print format, such an authoritative perspective not only serves to clarify technical aspects of the paper and to put it into a wider context, but also acts as an informative back-up to the press release, thereby promoting more responsible reporting by the media of controversial papers.

An example of a paper that notably needed simultaneous online publication of such a comment, given its politically loaded subject matter, is 'Genes mirror geography within Europe' (J. Novembre *et al. Nature* doi:10.1038/nature07331; 2008). This paper is remarkable, but are journalists likely to take the point that the two genetic components revealed by the authors represent just 0.45% of European genetic variability? That is, that 99.55% of the genetic variability is left aside in the figures and is not correlated with geography? Will they realize that the genetic maps were drawn from a subset of 1,387 in 3,192, from which individuals had been removed who had grandparental ancestry encompassing different regions?

The recent history of ethnic violence in Europe and elsewhere makes it doubtful whether the 'promise of ancestry tests' mentioned by Novembre and colleagues is likely to be a blessing for humanity. Such tests would be ideal tools for implementing discriminatory policies.

Vincent Detours Institute for Interdisciplinary Research, Université Libre de Bruxelles, Campus Erasme, CP602, 808 route de Lennik, 1070 Brussels, Belgium
e-mail: vdetours@ulb.ac.be

Last chance to save one of world's most species-rich regions

SIR — A visionary proposal from the small oil-exporting nation of Ecuador, to help combat climate change while protecting one of the most biodiverse spots on Earth, is under threat because of a lack of support.

The Yasuni-ITT initiative is the Ecuadorean government's limited-time offer to refrain from exploiting its largest untapped oilfield, in exchange for financial compensation from the international community. These oil fields lie beneath the core of Yasuni National Park, one of the most species-rich and intact parts of the Amazon. This remote region is also still home to isolated indigenous peoples.

We argue that people need to embrace such novel propositions if the world is to avoid increasing atmospheric carbon dioxide to catastrophic levels by burning the planet's remaining fossil fuels. Forgoing extraction of oil reserves in remote and/or sensitive places could be an important component of a larger plan to limit carbon emissions.

Ecuadorean president Rafael Correa launched the initiative in June 2007, asking the world to respond within one year. The deadline has been pushed back to December 2008, but this important proposal may be in danger of floundering because of a lack of committed financial contributions from the international community.

As the proposal enters a critical stage, we urge potential donors, including philanthropists, banks and governments, to consider the proposal carefully. An Ecuadorean council dedicated to the initiative could handle negotiations.

In June 2008, the German parliament gave the Yasuni-ITT initiative new life by formally backing it and calling on the German federal government to contribute financially. But more parties must follow this lead. Ecuador is seeking half of the projected revenues, or about US$350 million a year for 10 years.

Fair sharing of global responsibility will protect two of the world's most vital resources: the Amazon and our climate. For more information, please see http://yasuni-itt.info, or contact Lucia Gallardo at itt@mmrree.gov.ec.

Max Christian Namaste Solar, Boulder, Colorado 80302, USA
Matt Finer Save America's Forests, Washington DC 20003, USA
e-mail: matt@saveamericasforests.org
Carl Ross Save America's Forests, Washington DC 20003, USA

nature

AUTUMN BOOKS

Relics of the modern mind

Our enduring search for meaning in life explains the reverence with which the bones of seventeenth-century French philosopher René Descartes were worshipped, suggests **Lisa Jardine**.

Descartes' Bones: A Skeletal History of the Conflict Between Faith and Reason
by Russell Shorto
Doubleday: 2008. 320pp. $26

ILLUSTRATIONS BY S. KAMBAYASHI

In spring 1666, the body of the French philosopher and scientist René Descartes was removed from its resting place in the undistinguished Adolf Fredriks churchyard in the northeast of Stockholm where it had been buried quietly 16 years earlier, and transported under cover of darkness to the residence of the French ambassador to Sweden, Hugues de Terlon. There, under de Terlon's reverential supervision, the coffin was opened. Its contents were transferred to a small, specially made copper coffin to be transported to a more appropriate resting place in Paris. Thus began the curious peregrinations of Descartes' bones.

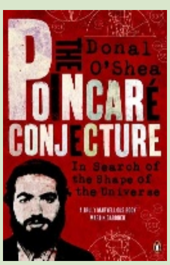
By the time de Terlon's convoy set off for home, Russell Shorto tells us in his beguiling new book, *Descartes' Bones*, fragments of the controversial thinker's earthly remains had already started to be dispersed, as admirers licitly and covertly acquired relics — souvenirs to be treasured and revered, and handed down through the generations.

Shorto confesses that when he began to investigate the story of Descartes' remains passing down through the centuries, he "fell in love with it, the way you can only fall in love with something truly odd that you find buried in a very old book". As he pursued the fortunes of the bones — and above all the separated skull — through the historical records, he realized that there was more to the story than mere curiosity. Here was "a historical detective story about the creation of the modern mind".

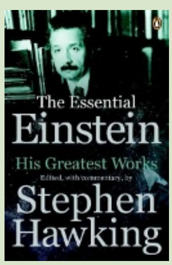
In their attitudes to the relics, Shorto argues, successive generations of Descartes' admirers and detractors betrayed their own deeply held beliefs about the limits of rationality, and the persisting need for faith in order to make sense of the complexity of human existence. In 1666, de Terlon had requested the authorities that he might himself be allowed, "religiously", to take possession of Descartes' right index finger, the bone "which had served as an instrument in the immortal writings of the deceased". The Swedish family that became the proud owners of Descartes' skull — how, it is not clear — had it lovingly inscribed with Latin verses celebrating its significance as a souvenir of the beginnings of rationality. Successive owners added their own signatures and inscriptions testifying to their own 'faith' in the relic.

According to the traditional history of philosophy, Descartes' statement "Cogito, ergo sum" and his committed view that the human body is a well-ordered machine, threatened to drive faith and human emotions out of the discourse of humanity. The story of Descartes' bones is, according to Shorto, the tale of each generation's struggle to retain its most deeply held spiritual beliefs in the face of such mechanistic explanations, an emotional attempt to bridge the mind–body dualism.

NEW IN PAPERBACK



The Poincaré Conjecture: In Search of the Shape of the Universe
by Donal O'Shea (Penguin, £9.99)
One of the great mathematical theories, the Poincaré conjecture defied proof for a century until it was solved in 2006 by an eccentric Russian mathematician. In his enthralling book, O'Shea explains the mathematics behind the conjecture, its proof and the curious personalities involved.



The Essential Einstein: His Greatest Works
Edited, with commentary, by Stephen Hawking
(Penguin, £12.99)
Collecting the highlights of his work, this book presents Einstein's view of physics and the world in his own words. Hawking introduces each section, summarizing the historical context and impacts of Einstein's discoveries.

Each step in the detective investigation contributes to our understanding of these same issues today. At the height of the French Revolution, for instance, Alexandre Lenoir became the self-appointed custodian of the human remains and historic relics preserved across France in churches and chapels, volunteering to rescue them from the desecration of the marauding mobs and place them in his own 'hall of antiquities' for safe keeping. He went about his task with revolutionary — or perhaps, Shorto suggests, religious — zeal, assembling huge quantities of stonework and statuary, and logging its acquisition with care.

As reason replaced faith in the new French Republic, Lenoir rescued the bones of France's greatest rationalist from the church of St Geneviève in Paris where they had lain since de Terlon recovered them from Sweden. Meanwhile, on the eve of the Terror, a French revolutionary committee decreed that the bones should be moved to the Pantheon — the new secular cathedral of the revolution — and a statue erected to Descartes' memory. Not for the first time, however, events overtook the best of intentions. In spite of an official decree of the ruling National Convention, the bones stayed in Lenoir's macabre museum, where he laid out his tombs and funerary statuary in a pleasure garden in which visitors could wander at will and "ponder beauty and death".

Once again, the whereabouts of Descartes' bones become shrouded in mystery. It is not even clear that Lenoir's rescued remains were those of the father of rationalism. Nor is it clear what happened to them thereafter. They seem to have gone missing among the carefully documented treasures in Lenoir's museum.

For Shorto, his own fascination with this curious piece of narrative history is a mirror for the concerns of each and every one of us, bewildered by modernity and struggling to find meaning and belief in a confusing world. His piecing together of the fortunes of Descartes' bones is elegantly interwoven with the intellectual history of Cartesian philosophy and beliefs, as well as with the history of the uncomfortable relationship between Catholicism and Cartesianism. His suggestion is, I think, that we cannot escape from our felt need for faith and devotion, and that, deprived of religious relics we turn to secular forms of worship.

I am not sure that our attitude to mortal remains has not always sat apart from the rest of our thinking and feeling — a place of talisman and taboo that brings out in us, from the depths of our soul as it were, our most primitive, elemental feelings and attitudes, however rational our outlook.

Between 1640 and 1660, for example, an English gentleman named John Bargrave embarked on a 'grand tour' of Europe in search of unusual items to assemble in a fashionable cabinet of curiosities. When he returned to England he became a canon of Canterbury Cathedral, to which he bequeathed his carefully catalogued collection. The Bargrave cabinets can still be seen there today. It is, in a way, a biography in a box, testifying in its varied items to Bargrave's personal passions and interests. There is a beautifully constructed model of an eye, which can be dismantled to show its inner structure. There are 34 saints' bones bought for 2 shillings and 6 pence (about £200 or US$350 today). There are tiny antique statues, rings, trinkets, each lodged in its own drawer and accompanied by a description of what it is and where it was found.

Just as in the case of Descartes' bones, Bargrave's cabinets reveal an intense fascination on Bargrave's part for relics and human remains. One of the items in it — which I have myself, somewhat tremulously, handled — is 'the finger of a Frenchman'. Bargrave's catalogue tells us of the "danse macabre" put on for him at the monastery at Toulouse where he acquired it and where "all the dead men and women's corpses that are buried there are dried into a kind of mummy". He wrote: "They showed us the corpse of a physician (of their acquaintance), which, when they put a clean piece of paper into one hand and a pen into the other, stood in such a posture as if he had seriously been a-writing a dose or prescription."

Bargrave's macabre trophy is a reminder of our enduring fascination with remnants and reminders of our own mortality. Shorto's beautifully written book does likewise — the reader will find its intellectual insights entertaining, enlightening and, perhaps, disturbing. ■

Lisa Jardine is director of the Centre for Editing Lives and Letters and centenary professor of Renaissance Studies at Queen Mary, University of London, London E1 4NS, UK. Her latest book is *Going Dutch: How England Plundered Holland's Glory*.

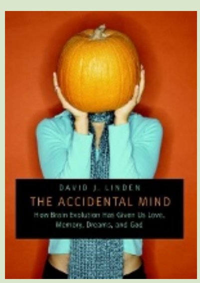
Psychology's pet subject

Alex & Me: How a Scientist and a Parrot Discovered a Hidden World of Animal Intelligence — and Formed a Deep Bond in the Process
by Irene Pepperberg
HarperCollins: 2008. 240 pp. $23.95

Alex and Me is the conjoined life story of cognitive psychologist Irene Pepperberg and her African grey parrot, Alex. Pepperberg sketches her life from lonely girlhood in Brooklyn, New York, through college and graduate school — she attended the Massachusetts Institute of Technology aged 16, and gained a PhD in theoretical chemistry at Harvard University — to her acquisition of Alex in 1977. Pepperberg helped Alex to develop the use of spoken words to identify numbers, colours, materials, shapes and categories of objects. She struggled to gain respect and recognition from the academic community for her work, and this takes centre stage in her account.

The book starts and ends with Alex's death in September 2007. The public response to his demise underscored his celebrity status. Outpourings of grief reached Pepperberg from all around the world — even *The Economist* ran an obituary. She conveys effectively her shock at losing not just her main research subject, but

New Theories of Everything
by John D. Barrow (Oxford Univ. Press, £8.99)
Updating his 1991 work, *Theories of Everything*, Barrow looks at the latest discoveries and theories about the Universe and explores their philosophical and cultural implications. He shows how the search for a Theory of Everything leaves us more aware of the richness and complexity of the cosmos.



The Accidental Mind: How Brain Evolution Has Given Us Love, Memory, Dreams, and God
by David J. Linden (Harvard Univ. Press, $17.95)
Linden's introduction to brain science is enjoyable and accessible. "Even readers trained in neuroscience are likely to enjoy the many tidbits of rarely taught information — on love, sex, gender, sleep and dreams — that spice up Linden's main argument," wrote reviewer Georg Striedter (*Nature* **447,** 640; 2007).

also a beloved friend of some 30 years.

Similarly to many scientists whose lives revolve largely around their work, Pepperberg is impatient with her critics. She bemoans the reviewers of her work who force her to carry out more experiments when the results are already abundantly clear, and the funding agencies that fail to see the importance of her project. She calls their criticisms "denigrating". Yet there are many good reasons to be sceptical of the exceptional abilities of a single animal.

More than 350 years ago, the philosopher René Descartes suggested that there was a fundamental difference between "teach[ing] a magpie to say good-day to its mistress, when it sees her approach" and the verbal expression of ideas by human beings. The magpie, declared Descartes, does not form words as expressions of its thoughts, only as responses to obtain a desirable reward, learnt through training.

Pepperberg insists that her work with Alex puts the final nail in Descartes' coffin, and shoos away the criticism of the behaviourists who are Descartes' heirs. She quotes her last interchange with Alex on the night before his death: "'You be good,' said Alex. 'I love you.'"

Although she admits that her students would sometimes teach Alex entertaining sayings in the manner that Descartes suggested, she insists that her relationship with Alex was truly one of reciprocated affection. However, before concluding that a parrot proclaiming love knows whereof he squawks, one would want to be certain that nobody has been giving the bird a titbit for what he says.

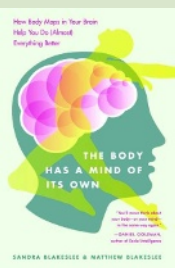
A century ago in Berlin, a respected school teacher presented to the public a horse known as Clever Hans, who had the apparent intelligence of a 13-year-old child. A prestigious commission of experts was convinced that Hans could solve school-level arithmetic problems. However, psychology student Oskar Pfungst demonstrated that the horse was not answering the questions as posed, but instead responding to the subtle head movements of those who asked the questions. The movements were so slight the questioners did not know they were making them, and there was no suggestion that the horse's owner had intentionally taught Hans to respond to them. Can we be confident that Alex was not deducing the correct response from subtle cues the experimenters were giving off?

One defence against the accusation that Alex was just another Clever Hans is that his responses were much more varied than the foot stomping that Hans used to answer questions. Whereas Hans only needed to be cued to start and stop moving his hoof, Alex needed to say many different things to answer a question correctly. However, the fact that Pepperberg always remained in the room to interpret Alex's utterances poses problems for this defence. We know that the human ear is easily tricked. Led Zeppelin's *Stairway to Heaven* contains satanic lyrics when played backwards — but only if the words are shown to the listener. Without a script, the reversed sounds are perceived as nothing more than mixed-up phonemes. The squawking of a parrot often contains jumbled speech sounds, so the form these take to a listener might well depend on what that person expects to hear.

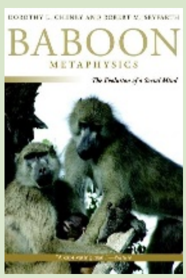
Alex and Me is an engaging narrative because it has the intimate relationship between one human and one parrot at its centre. But as an exploration in science, it is deeply worrying that nobody can replicate its central findings, given the ease with which the subjects and equipment can be acquired. Even Pepperberg has been unable to replicate Alex's achievements using other parrots.

Alex may have had some sense of number, shape, colour and material. Similar results have been obtained with pigeons, so it is not an absurd conjecture. For those who did not know him personally, the tragedy of Alex's passing is that the records that remain are not enough to prove the case one way or the other. ■

Clive Wynne is associate professor of psychology at the University of Florida, Gainesville, Florida 32601, USA, and author of *Do Animals Think?*
e-mail: wynne@ufl.edu



The Body Has a Mind of Its Own: How the Body Maps in Your Brain Help You Do (Almost) Everything Better
by Sandra Blakeslee and Matthew Blakeslee
(Random House, $15.00)
This book explores the emerging science of mind–body connections. "Ideas from sister disciplines, including neurology, psychiatry and cultural anthropology, mix comfortably with laboratory observations," wrote Edvard Moser (*Nature* **449,** 406; 2007).



Baboon Metaphysics: The Evolution of a Social Mind
by Dorothy L. Cheyney and Robert M. Seyfarth
(Univ. Chicago Press, $18.00)
Cheyney and Seyfarth examine the dynamics of baboon society. "Their enthusiasm is obvious, and their knowledge is vast and expressed with great clarity," wrote Asif A. Ghazanfar. "It will get you thinking — and maybe spur you to travel to Africa to see it all for yourself." (*Nature* **448,** 535–536; 2007).

How a hoax fooled the faithful

The Sun and the Moon: The Remarkable True Account of Hoaxers, Showmen, Dueling Journalists, and Lunar Man-Bats in Nineteenth-Century New York
by Matthew Goodman
Basic Books: 2008. 384 pp. $26.00, £15.99

In the hot summer of 1835, Matthew Goodman writes in his new book *The Sun and the Moon*, an English-educated journalist who had moved to New York mulled over his adopted country. Although there was much that was admirable about the nation, he decided, a blind obedience to religion was keeping it from being as rational as he wished it to be. Being a superior sort of individual, and confident that he was better educated than those around him, he resolved to fix that.

The problem was most painful, he felt, in astronomy, which was especially popular that year. Halley's comet was expected back in September, and already in August telescopes had been set up in public parks to offer citizens a first glimpse for 6 cents. Yet, instead of the predictable nature of the comet's return demolishing religious obedience, as popular science writer Richard Dawkins might have wished, public opinion was just the opposite.

Halley's comet, it was believed, proved the existence of a divine Creator, who was able to propel objects at will through the heavens. New Yorkers — educated and not — accepted that intelligent beings existed on other planets, providing further proof that a divine creator beneficently ruled over us all.

That link between astronomy and religion was what Richard Adams Locke, a descendant of the philosopher John Locke, decided had to be changed. He knew that the founding fathers of the United States had been critical of revealed religion — indeed, during his own education in Britain he had come to agree with that criticism. Before his arrival in New York he had looked to the United States as a beacon of radical reform, suitable to inspire class-bound, subservient Britain. Yet now his adopted country was in the grip of religious fundamentalism.

After several years of making his way in journalism, Locke was in a position to act. He had recently been promoted from police reporter to editor-in-chief of the *New York Sun*, one of the world's first mass-market newspapers. It was aimed at ordinary readers, produced on new, steam-driven presses, and sold not by genteel subscription but by hustling newspaper boys.

Late in August 1835, a series of articles began appearing in the newspaper. They described the wondrous discoveries of John Herschel, the noted English astronomer who had gone to the Cape provinces, in what is now South Africa, to chart the southern sky. Herschel was indeed there, but being so remote he was also many weeks away from any direct communication with New York. This made him an especially good subject for an inventive journalist. The hoax that Locke conjured up would have a long stretch to run before it could be disproved.

The first few *Sun* articles were innocuous enough, stating that a scientific journal was reporting some astounding results before going into technical detail about the improved telescope that Herschel was using.

Then came the bombshell: Herschel, it was reported, had discovered lakes of water and herds of wild animals on the Moon. Circulation of the *Sun* shot up. More reports came in. Herschel had discovered more than animals — he had identified human-like life there as well.

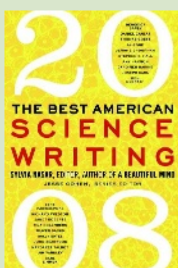
The newsboys selling the *Sun* might not have read the articles in detail, but they loved the circulation they generated. Everyone — clerics and engineers, government officials and labourers — wanted to know what was going to come next.

Locke didn't disappoint them. For shortly before the great southern African telescope needed to be, conveniently, repositioned for other pressing enquiries, Herschel managed to detect clear marks of religious temples on the Moon.

By Locke's plan, that would end it. He had taken the purported discoveries to ridiculous extremes. Everyone would realize he was writing a spoof, and Americans would see how silly it was to use astronomy as a way of justifying their religious beliefs.

But English journalists who think they know the United States have a habit of getting things wrong. Even when reports came back

Bad Science
by Ben Goldacre
(HarperCollins, £12.99)
Fast-paced and highly readable, *Bad Science* dismantles the dubious claims behind scare headlines, fad nutritionists and suspect 'clinical trials'. Looking at specific cases, Ben Goldacre explains how bad science is presented to look real and why people are so willing to believe it.



The Best American Science Writing 2008
Edited by Sylvia Nasar (series edited by Jesse Cohen) (Harper Perennial, $14.95)
The latest edition of this popular series offers a selection of excerpts from familiar names, including Oliver Sacks and Carl Zimmer. Some focus on climate change or new technology. Many deal with medicine or disease — from genetic disorders, psychiatry and cancer, to compromised physicians and misinterpreted research.

from South Africa that Herschel had made no such discoveries as the *Sun* described, good New Yorkers were only reinforced in their religious beliefs. Surely the visions that he had described were a sign of the religious insights that would await us, once science moved forward enough?

The Sun and the Moon is a wonderful cautionary tale, especially in an era like our own, when, for example, no candidate is likely to be elected president of the United States if he professes disbelief in either science or religion. The mix seems illogical, but only if one misses how the two domains are popularly understood. Most people do not understand the details of how science works. They just see the impressive results of what it can achieve — whether it is steam engines in Locke's era or iPhones in ours. The more technology advances, the more such blind trust is encouraged. Locke and Dawkins might well weep, but a popular life that shares none of their ultimate views goes on.

Unfortunately, Goodman tells the story poorly, undecided if his book is an account of Locke's hoax or a general social history of the mass press and ways that cynics and reformers dealt with popular credulity. But his research is excellent, and if publishers fund him to write a wider account, it would be well worth reading. ■

David Bodanis is a writer based in London and the author of *Passionate Minds*.

Tale of a clockwork computer

Decoding the Heavens: Solving the Mystery of the World's First Computer
by Jo Marchant
William Heinemann/Da Capo Press: 2008/2009. 330 pp/288 pp. £12.99/$25

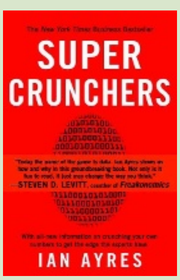
It is appropriate that Arthur C. Clarke recommended this book before his death in March 2008. *Decoding the Heavens* tells the story of the 2,000-year-old Antikythera Mechanism and its decipherment. The tale encompasses diving, shipwrecked treasure, scientific puzzles, cutting-edge technology and computing, all of which Clarke knew from the inside. Moreover, he had a walk-on role.

The Antikythera Mechanism is an ancient Greek artefact comprising more than 30 precisely cut bronze gear wheels, dials and pointers held in a wooden case.

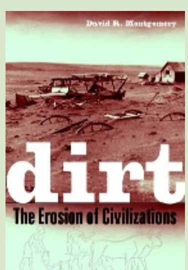
Clarke began to publicize the mechanism in 1959, perhaps encouraged by his friend Jacques Cousteau, who in 1953 had reinvestigated an ancient shipwreck off the Greek island of Antikythera, which had been salvaged by roving Aegean sponge divers in 1900–01. In 1965, Clarke tracked down the mechanism's encrusted remains languishing in a cigar box at the National Archaeological Museum in Athens, later televising it in *Arthur C. Clarke's Mysterious World* in 1980. If the ancient Greeks had been able to build on the knowledge required to make the mechanism, he told a Smithsonian Institution audience in 1973, by now humans might be exploring the nearer stars, not merely the inner Solar System.

Yet, as Clarke admits on the book's jacket, the Antikythera Mechanism remains "underrated". The instrument was known only to a small circle of enthusiasts until a paper was published in 2006 explaining its operation, based on X-ray images of the interior and a new technique for illuminating previously hidden surface inscriptions (T. Freeth *et al. Nature* **444,** 587–559; 2006). Jo Marchant, then a *Nature* staff journalist, wrote a news feature to accompany the paper, and the book grew from her research. *Decoding the Heavens* is Marchant's first book and is aimed squarely at the non-scientific reader, but is serious about the science. Her gripping and varied account will propel the mechanism to greater fame, although it may never achieve the celebrity of the Rosetta Stone that it probably deserves.

The complexity of the mechanism suggests that the ancient Greeks had achieved a level of clockwork technology not reached again until the Industrial Revolution of the eighteenth century. "It's hard to overestimate the uniqueness of the find. Before the Antikythera Mechanism, not one gearwheel had ever been found from antiquity, nor indeed any example of an accurate pointer or scale. Apart from the Antikythera Mechanism, they still haven't," Marchant explains.



Super Crunchers
by Ian Ayres
(Bantam Books, $14.00)
Ayres explains how data mining, statistics and number crunching are used increasingly by everyone and everything from Internet search engines to your family physician. Showing how this affects people directly, he gives advice on how to defend ourselves against the negative sides of this data blitz.



Dirt: The Erosion of Civilizations
by David Montgomery
(Univ. California Press, $16.95)
Montgomery explains how the collapse of ancient civilizations is linked to soil erosion and shows that this can be seen today. "Societies prosper and persist best when they figure out ways to keep their soil where it belongs and not treat it as if it were dirt cheap," wrote Eric A. Davidson (*Nature* **447,** 777–778; 2007).

The original 1900–01 dive was the first archaeological exploration of a wreck, undertaken decades before Cousteau's invention of the scuba apparatus. After describing its drama, Marchant turns to the scholars from many countries who tried to date and interpret the mechanism over the next century. In the 1960s and 1970s, the leading investigator was the late science historian Derek de Solla Price of Yale University, who published *Gears from the Greeks* (American Philosophical Society, 1974). Since the 1990s, two rival explanations have emerged: one from Michael Wright, a former curator at the Science Museum in London, who is inspired by de Solla Price, and the other from the multidisciplinary team that published in *Nature*, led by mathematician and film-maker Tony Freeth and Mike Edmunds, an astronomer at Cardiff University. Marchant's handling of the rivalry is excitingly tense.

Damaged Greek inscriptions initially dated the mechanism to between the second century BC and the second century AD. Radiocarbon analysis dated the ship's wooden hull to between 260 and 180 BC, although the ship could have sailed later. The design of the wreck's pottery amphorae narrowed the date to the first half of the first century BC. Silver and bronze coins found by a second Cousteau expedition in 1976 showed the ship sank between 70 and 60 BC, probably while sailing from Pergamon, on the coast of Asia Minor, towards Rome.

The mechanism was initially thought to be an astrolabe. But its gear wheels suggested another use, as astrolabes do not need gears. De Solla Price imagined it was a calendar computer, but this did not explain the ratios of teeth on the gears. Wright now favours a planetarium explanation, whereas Freeth's team believes it is an instrument for predicting eclipses. Freeth prefers to call it a 'calculator' rather than a 'computer', whereas Marchant, following historian Doron Swade's definition that a computer can display its calculated numbers on a scale, uses the latter. In both explanations, the mechanism's builders assumed that Earth is at the centre of the heavens, as believed by Greek astronomers and philosophers except for the heliocentric Aristarchus of Samos, and that the movement of the heavenly bodies is based on epicycles.

Marchant's fascinating final chapter asks who made it, and why. The inventions of Archimedes in the third century BC might have influenced it; an epigraphic analysis published earlier this year hints at a link with his home city of Syracuse (T. Freeth *et al. Nature* **454,** 614–617; 2008). The astronomer Hipparchus may also have inspired its design, but was probably dead before its construction. The philosopher Posidonius, who lived at the right time, is a strong candidate; his student Cicero mentions that Posidonius built an instrument that reproduced the "motions of the Sun, the Moon and the five planets".

The mechanism's purpose may have been for astrology, in which Hipparchus apparently believed, but more likely it had no practical use. Like the orreries of the Enlightenment, it was probably one among many luxury gadgets for the educated Greek and Roman elite — the only example of its kind to survive. "The Antikythera Mechanism," Marchant concludes, "was originally meant as a celebration of the heavens." ■

Andrew Robinson is a visiting fellow at Wolfson College, Cambridge. He is the author of *Lost Languages: The Enigma of the World's Undeciphered Scripts* and *The Story of Measurement*.
e-mail: ar471@cam.ac.uk

Liver satire delivers

Liver: A Fictional Organ with a Surface Anatomy of Four Lobes
by Will Self
Viking: 2008. 288 pp. £18.99

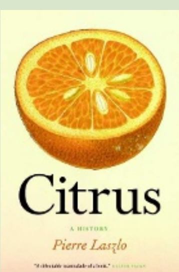
We live, by and large, Panglossian lives, naively reading purpose into the world, and into ourselves. Confronted by our bodies, we wonder at their watch-like precision. Each part of us does something. A heart pumps blood. A skull shields the brain. Eyes see. Hands grip.

When we look at the liver, however, we see not a tool for living, but a living thing — and this is an uncomfortable proposition. The liver has a plethora of functions: storage of the carbohydrate glycogen, decomposition of red blood cells, production of blood plasma proteins, detoxification, bile production — and that's without even mentioning its regulatory abilities. It intimidates us by its efficient ubiquity.

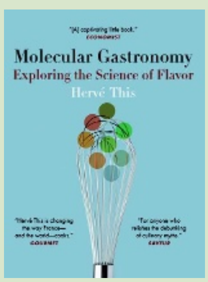
Will Self's new book, *Liver*, is not body horror in the science-fiction sense. For Self, as aficionados of the author will expect, *Liver* is satire. In his vision, our livers are more valuable than we are, more able, more alive. The liver is the only internal organ in the body that can regenerate itself to a significant extent. Yet still we contrive, over the course of our lives, to squander its magnificent estate.

The four 'lobes' of Self's book are individual stories — his strongest in years. With a little intertwining of narrative, his peculiar tales of abuse, disease and decay largely follow their own paths. Cirrhosis, cancer, hepatitis and, with a nod to Greek mythology, a vulture wander among the human protagonists as equals. Self's satire is classical rather than radical, rooted more in Alexander Pope than Jonathan Swift. Diseased flesh takes plenty of collateral damage, but contemporary behaviour is Self's real target. "Confronted with the nobility of feeling, high culture and deep spirituality", Self — like the malevolent bar fly of his first story, 'Foie Humain' — "sees nothing but the stereotypic behaviours of anthropoid geese."

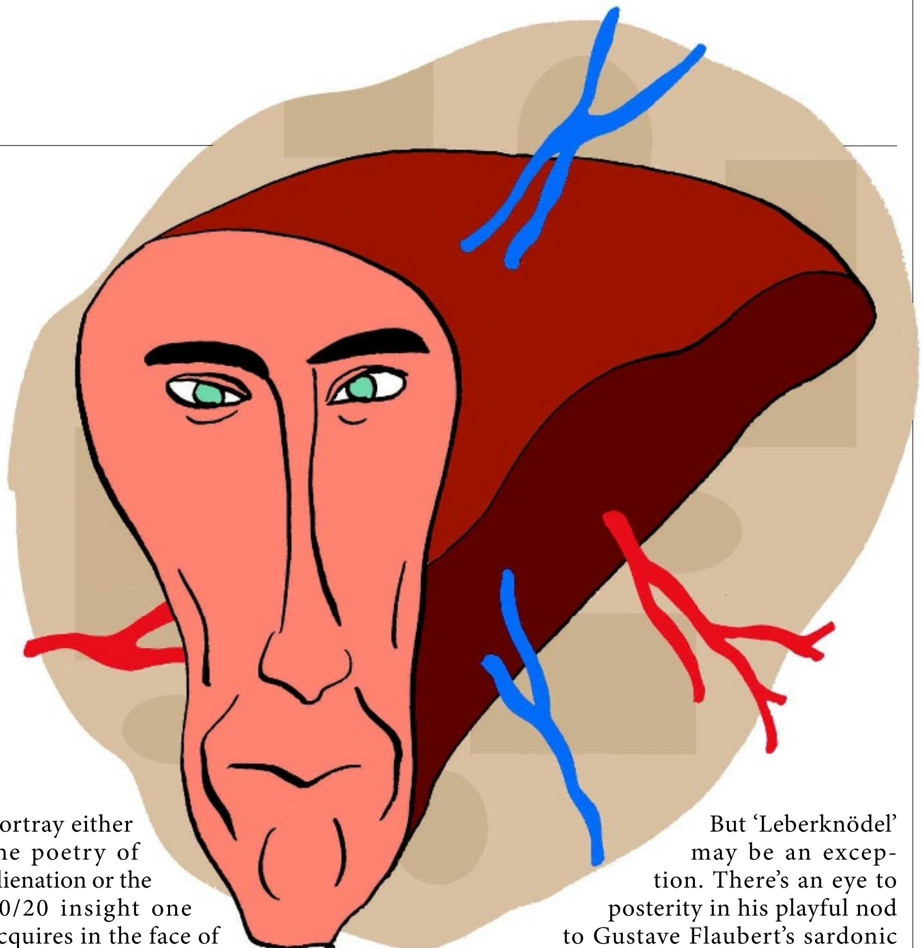
Satire depends for its success on a pitiless accuracy. Self's prose veers between the appetizing and the nauseating, yet it is almost always on the nail: proof that the more accurately you describe a thing, the more surreal it seems. His vivid characters include a cancer sufferer being hustled towards her elective suicide by her daughter's poor timekeeping, the determined alcoholic gorging of a hapless barman, and a dangerously unmotivated Harley Street hepatologist. These sketches



Citrus: A History
by Pierre Laszlo (Univ. Chicago Press, $17.00)
Laszlo describes the chemistry of citrus fruits. "His succinct explanations of the Maillard and caramelization reactions when describing how to make marmalade, and his notes on foams, phase separation, and the volatility of aromas when explaining how to make a sabayon, justify the cost of the book on their own," wrote Peter Barham (*Nature* **450,** 479; 2007).



Molecular Gastronomy: Exploring the Science of Flavor
by Hervé This (Translated by Malcolm DeBevoise) (Columbia Univ. Press, $19.95)
French experimental chef, This, opens up the kitchen to scientific experiments. In this enjoyable book, he debunks old rules and superstitions about cooking, and shows how food can be looked at, prepared, cooked and tasted in new ways.



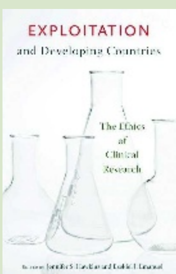
portray either the poetry of alienation or the 20/20 insight one acquires in the face of approaching death.

Self plays both sides, nowhere more affectingly than in the collection's magnificent centrepiece, 'Leberknödel' (Liver Dumplings). Joyce, a retired hospital administrator, knew that her cancers would not stop "until they had toppled the sovereignty of consciousness itself, and replaced it with their own screaming masses of cancerous tissue". Appalled at the 'bad habit' of terminal decline, she arranges her own death. Seemingly reprieved, she finds, however, that her living has become as much of a bad habit as her dying. She looks objectively at her life — her world, her friends, her sot of a daughter — and realizes she has lost the empathic ability that gave her life its meaning. She has, essentially, already killed herself.

Self is not a writer who worries about the long-term value of his art, as that would distract him from the ephemera so essential to it. But 'Leberknödel' may be an exception. There's an eye to posterity in his playful nod to Gustave Flaubert's sardonic prose technique, gleefully italicizing every cliché passing through Joyce's dying mind. Like Madame Bovary, 'Leberknödel' mixes its scorn with warmth and sympathy. As a result, one gets a sense of the author's moral authority. Our bodies are not ours; nor are our feelings. We think our perceptions are ephemeral, but they are rooted in a physics that will outlast us. Joyce manages to argue herself away, destroyed by her own strength of purpose. This vision of humanity — too strong for itself, too clever by far — is more tragic than satirical.

For Self to kick the chair out from under himself in this way, turning his trademark scorn to tears of sympathy, is an achievement. *Liver*'s grotesques are not meant to hold — but Joyce endures. ■

Simon Ings is a writer based in London. His most recent book is *A Natural History of Seeing*.
e-mail: simon@fisheye.demon.co.uk



Exploitation and Developing Countries: The Ethics of Clinical Research
Edited by Jennifer S. Hawkins and Ezekiel J. Emanuel
(Princeton Univ. Press, $14.95)
Scientists and philosophers delve into the moral implications of clinical research in developing countries. Different chapters examine case studies and look at the benefits and costs of clinical trials, both to the individual subjects and the different nations involved.

Drug-fuelled counterculture

Albion Dreaming: A Popular History of LSD in Britain
by Andy Roberts
Marshall Cavendish: 2008. 288 pp. £18.99

LSD — lysergic acid diethylamide — is an evocative acronym: it strikes a rational fear into governments and an irrational one into the media. Those who have taken the drug are in awe of its transformative power; to those who have not, it is a dangerous unknown.

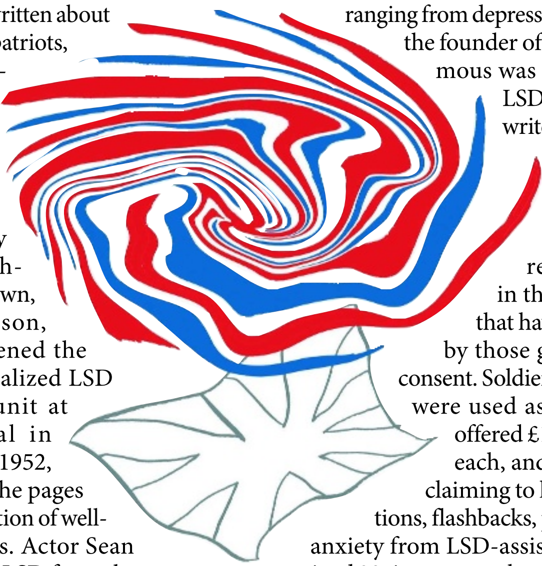
Since LSD was invented by Swiss chemist Albert Hofmann, who referred to it as his "problem child", it has been a drug in search of a use. LSD is now inching its way back onto research agendas in the United States and Europe.

Albion Dreaming is a timely book, especially given Hofmann's death this year. It examines how a compound viewed by some as having unmatched potential for positively transforming society through individual enlightenment, instead fuelled a decades-long counterculture, which was eventually suppressed. With investigative zeal, Andy Roberts charts the highs and lows of LSD's relationship with the anarchists and intellectuals of 1950s, 1960s and 1970s Britain, and takes the reader on a magical mystery tour through the cultural changes catalysed by the most potent chemical known to humankind.

The British government was at first intrigued by LSD's potential. When the compound reached its shores in 1952, the medical and military establishments were keen to investigate and exploit the drug in psychotherapy and as a battlefield weapon, respectively, as was the Central Intelligence Agency across the Atlantic.

The United States is rightly considered to be the cheerleader of LSD-fuelled counterculture, but the author convincingly demonstrates that Britain had a strong supporting role. The term 'psychedelic' was invented in a creative wordplay between two Englishmen, the writer Aldous Huxley and his doctor friend Humphrey Osmond, who first gave him mescaline. Huxley's "To make this trivial word sublime, take half a gramme of phanerothyme [soul-visible]", was topped by Osmond's "To fathom hell, or soar angelic, take a pinch of psychedelic [mind-manifesting]."

It was a Briton too, Michael Hollingshead, who turned Harvard psychologist Timothy Leary on to LSD, after which Leary morphed into a self-styled psychedelic guru and urged the US youth to "Turn on, tune in, drop out", which they did in their millions.



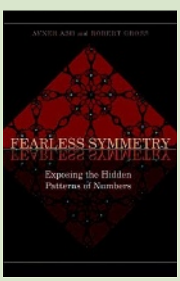
Much has been written about Leary and his compatriots, but *Albion Dreaming* retains a gritty UK perspective. From the still-secret experiments conducted at Britain's military research establishment at Porton Down, to Ronald Sandison, a doctor who opened the world's first specialized LSD psychotherapy unit at Powick Hospital in Worcestershire in 1952, the author packs the pages with a riotous selection of well-sourced anecdotes. Actor Sean Connery sampled LSD from the psychiatrist's couch and his therapist demanded a fine single-malt whisky as part payment. The BBC's laudable — or is that laughable? — attempts to filter out 'drug-inspired' lyrics on radio led to a ban on The Beatles' song *A Day in the Life*, but *Lucy in the Sky with Diamonds* slipped through the net. Corrupt Drug Squad officers paid informants with LSD that had been seized in earlier raids, until the well-organized Operation Julie smashed a major international manufacturing ring later in the 1970s, when up to half of the world's illicit LSD was made in Britain.

Although Roberts does not match the gripping narrative of Jay Stevens' *Storming Heaven: LSD and the American Dream* (Atlantic Monthly Press, 1987), he spins a good yarn, distancing rumours such as Francis Crick's links to LSD from established fact. He found no proof to back the claim that Crick took LSD and saw the double helix in his visions. Conspiracy theories, namely allegations of the media's complicity with the police or the secret intelligence services, are also maturely circumvented.

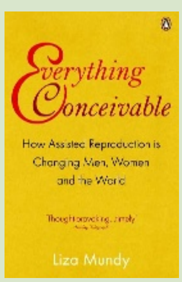
The book could have included more about the scientific applications of LSD. Thousands of papers in the 1950s and 1960s suggested that LSD could be used as a treatment for conditions ranging from depression to alcoholism — the founder of Alcoholics Anonymous was transformed by an LSD trip — but Roberts writes a popular history, not a scholarly one. *Albion Dreaming* provides the fullest account yet of recent legal battles in the United Kingdom that have been resurrected by those given LSD without consent. Soldiers who claimed they were used as guinea pigs were offered £10,000 (US$18,000) each, and nearly 50 patients claiming to have had hallucinations, flashbacks, paranoia or chronic anxiety from LSD-assisted psychotherapy gained 20 times as much compensation.

The street-level perspective of the book works, but an over-reliance on the exploits of individuals leads to a lack of deeper insights. The concluding pages are wasted on an intelligent but irrelevant argument for the legalization of drugs, yet I wanted to know from the author why research on LSD has yet to resume in the United Kingdom, or why the psychedelic-powered revolutions did not take off across the world. Consequently, *Albion Dreaming* is an enjoyable journey but leaves the reader with more questions than answers — much like the LSD trip itself. ■

Arran Frood is a science writer based in the United Kingdom.



Fearless Symmetry: Exposing the Hidden Patterns of Numbers
by Avner Ash and Robert Gross
(Princeton Univ. Press, $18.95)
This book attempts to explain complex maths in an interesting way. Reviewing the hardback edition, Timothy Gowers wrote: "The authors are to be admired for taking a very difficult topic and making it, if not fully accessible, then certainly more accessible than it was before." (*Nature* **446,** 26; 2007).



Everything Conceivable: How Assisted Reproduction is Changing Men, Women and the World
by Liza Mundy (Penguin, £9.99)
Mundy looks at the consequences of the "reproductive revolution" in science that is changing who reproduces and how. She covers changes in society and technology, talking to physicians, scientists, parents and campaigners.

A scientist's life for me

Forty years after the publication of James Watson's *The Double Helix*, Georgina Ferry asks why the life stories of so few scientists make it into the bookshops.

In 1968 Peter Medawar, Nobel prizewinner and author of many witty reflections on science and its practitioners, consented to write a preface to Ronald Clark's biography of the influential British biologist J. B. S. Haldane. Imagine Clark's consternation when he read its opening line: "The lives of academics, considered as Lives, almost always make dull reading." Later, Medawar recycled the opening paragraph for an essay in his collection *Pluto's Republic* (1982), claiming further that scientists' lives, unlike those of "artists and men of letters", were "not a source of cultural insight".

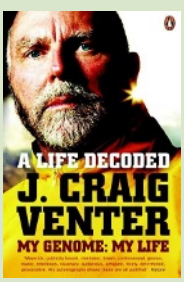
James Watson's *The Double Helix*, a book that broke the mould of scientific life-writing, also appeared in 1968. It provided abundant 'cultural insight' into the combination of good contacts, brilliance, luck, hard work and ruthless competitiveness that brought to light the DNA structure. It was panned by many of Watson's contemporaries — if Francis Crick had got his way, the book would never have been published. Yet in his own memoir *What Mad Pursuit* (1988), Crick later admitted that he was wrong: "I now appreciate how skilful Jim was, not only in making the book read like a detective story, but also by managing to include a surprising amount of science."

In *A Life Decoded*, published last year, genomic entrepreneur J. Craig Venter recalls choosing *The Double Helix* for a college assignment. Intending to train as a doctor on his return from military service in Vietnam, he sensed from its pages the adrenaline thrill of research. Generations of scientists have pointed to books that opened up this new world to them, notably the 1926 classic *The Microbe Hunters* by Paul de Kruif.

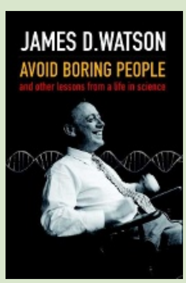
Where should today's bright 16-year-old, ambitious graduate student or interested general reader look for a personal insight into modern laboratory life? Scientific life-writing is now a small and shrinking enterprise. Publishers agree that the market for scholarly biography has suffered from the onslaught of celebrity memoir, with ghost-written autobiographies of models, sporting heroes and television personalities enjoying sales that a serious study cannot hope to match.

Selling biographies of scientists is particularly difficult, thinks Jenny Uglow, author of *The Lunar Men* (2002) and *Nature's Engraver* (2006), and editor at Chatto and Windus. "People look for subjects that are close to their interests, and they perceive scientists as just not of their world," she says. Booksellers face the same problem. Their choices of books to stock, now largely in the hands of a select few in the head offices of the major chains, influence the publishers. Most people have heard of very few scientists, so those that they do recognize — Isaac Newton, Charles Darwin and Albert Einstein — seem the safest bets. The rankings of online retailer Amazon show that of the top 100 scientific biographies in the United Kingdom, ten are by or about Darwin; in Germany, the first five places are all occupied by Einstein, whereas Richard Feynman and Benjamin Franklin feature strongly in the United States. None comes close to making the top 100 biographies.

Biography shelves in bookshops are well stocked with volumes about recently deceased or living figures in politics and the arts. A handful of scientists are represented: Venter's memoir joins well-received biographies of the crystallographer J. D. Bernal by Andrew Brown (2005)



A Life Decoded: My Genome: My Life
by J. Craig Venter (Penguin, £9.99)
In his review of the hardback edition, Jan Witkowski wrote: "Four decades on, our infinitely more vulgar media has called Venter many things: maverick, publicity hound, risk-taker, brash, controversial, genius, manic, rebellious, visionary, audacious, arrogant, feisty, determined, provocative. His autobiography shows that they are all justified." (*Nature* **449,** 785–786; 2007).



Avoid Boring People: And Other Lessons From a Life in Science
by James Watson (Oxford Univ. Press, £9.99)
Watson's frank autobiography covers personal areas as well as the discovery of DNA and his later controversies. It "weaves a deliciously detailed account of his life both in and out of science with a series of lessons drawn from those experiences," wrote Huntington F. Willard (*Nature* **449,** 787; 2007).

and of Francis Crick by Matt Ridley (2006), and Eric Kandel's autobiography *In Search of Memory* (2006) sold well in the United States. These books are few in number, increasingly the preserve of academic rather than trade publishers, and often relegated from the store's biography section to the obscure corner labelled 'science'.

It is regrettable that readers are not more alive to the cultural significance of scientists. But who maintains this gulf? Scientists publish their work in places where only other scientists will read it, in language that only other scientists understand. Many fear that attracting public attention will lose them the respect of their colleagues; others dismiss tales of scientific rivalry as gossip. Some argue that individuals are irrelevant to the progress of science: anyone could have discovered the double helix, but only Leonardo da Vinci could have painted the *Mona Lisa*.

In practice, science is done by real people, as different from one another — except in their devotion to their field of study — as any other sector of humanity. Telling their stories transforms the stereotype of the scientist into vivid individuality. Truthful biographies scotch the myth of the solitary genius: any contemporary scientist's story is threaded through the network of exchanges and rivalries within and between labs that makes it such an intensely social activity. Parents, children, spouses and lovers feature as prominently in the lives and careers of scientists as in those of artists.

Little hint of personal background emerges in the memoirs of deceased fellows written by close colleagues and published by scientific academies. These essays are useful research tools for historians, but not widely read by outsiders. Occasionally a scientist, having achieved some level of distinction, decides that it might be worth documenting his or her experiences for the public. But most have no name recognition and therefore no market beyond their immediate academic circle. Even within science, students learn so little of the history of their subjects that they have no real sense of the people behind the classic references they cite.

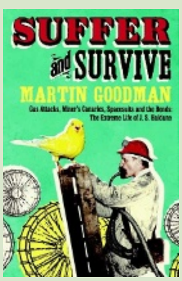
Biography and autobiography, it seems to me, offer an ideal opportunity to engage people's natural curiosity about the lives of others and so draw them into the quest to understand the physical world. Name recognition remains a barrier, but there are some encouraging pointers. Palaeontologist Richard Fortey has written a succession of brilliantly accessible personal reflections on his field of study and London's Natural History Museum. He admits that nobody had heard of him until he started writing popular books. For less gifted communicators, success depends on giving the job to a writer with an already formidable reputation. Simon Winchester is doing well with his new biography of the biochemist and Sinophile Joseph Needham (*The Man Who Loved China*, entitled *Bomb, Book and Compass* in the United Kingdom), as will Samuel Taylor Coleridge's celebrated biographer Richard Holmes, who has chosen Joseph Banks, Humphry Davy and William Herschel as subjects for his first major work in ten years, *The Age of Wonder*. The television tie-in works for science too, with Simon Flynn of Icon Books reporting that the BBC4 series that accompanied Piers Bizony's strongly biographical *Atom* (2007) "made a huge difference" to sales.

Alternatives to the cradle-to-grave biography or memoir are to be welcomed. The Internet is ripe for exploitation, and is a much more likely port of call for the young and impressionable. Short autobiographies of Nobel prizewinners are available on http://nobelprize.org. But any scientist can include on their web page or in their blog the story of how they got into science, who inspired them, and the joys and frustrations of their working lives.

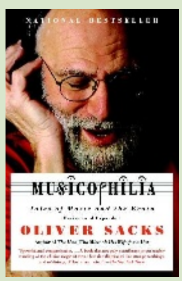
The medium is not the issue, however. Presenting the reality of the scientific life depends simply on the willingness of scientists to appear as individuals — and of their colleagues to applaud their doing so. ■

Georgina Ferry is a writer based in Oxford, UK. Her most recent biography is *Max Perutz and the Secret of Life*.

Discuss science biography online at http://tinyurl.com/3j6y44.



Suffer and Survive: Gas Attacks, Miners' Canaries, Spacesuits and the Bends: The Extreme Life of J. S. Haldane
by Martin Goodman (Pocket Books, £8.99)
Physiologist John Scott Haldane deliberately exposed himself to toxic gases. His results saved many lives, from mine-workers to deep-sea divers. Reviewer Andy Meharg commented: "It is a fitting tribute to a pioneer who enabled the human body to survive at the extremes of modern life." (*Nature* **449,** 981; 2007).



Musicophilia: Tales of Music and the Brain
by Oliver Sacks (Random House, $14.95)
Neurologist and medical writer Sacks delves into the world of music. Reviewing the hardback edition, Laura Garwin wrote: "Sacks is the consummate storyteller, and his extensive network of patients, friends and correspondents — supplemented by a magpie-like erudition — keeps him well supplied with raw material." (*Nature* **449,** 977–978; 2007).

ESSAY

India's rise to the Moon

Why does a developing nation have such an ambitious space programme? **Subhadra Menon** traces its foundations back to the work of one visionary physicist 60 years ago.

The space giants of the world are a growing family. This month, India plans to launch its first mission to the Moon: Chandrayaan-1 — Sanskrit for 'mooncraft' — is scheduled to blast off from the island of Sriharikota in the Bay of Bengal between 22 and 26 October. The craft is intended to orbit the Moon for two years, send an impactor to the surface and use remote sensing to produce chemical and topographical maps.

India's space programme has grown against the odds. A few decades ago, it struggled against crippling sanctions imposed by the United States after it exploded a nuclear device in 1974. Yet today, the Indian Remote Sensing satellite system, with its seven active satellites, is the largest family of remote sensing satellites in the world for civilian use. And the country's 11 national communications satellites form the largest such constellation in the Asia–Pacific region. This month's unmanned mission to the Moon marks a leap into more ambitious science-based projects, and the beginning of a string of such missions for the country.

A burst of enthusiasm about space in recent decades has led some to talk about an Asian space race. The motivations of these new players are remarkably similar to those that spurred the first space race: self-reliance, national pride, security and economic gain. India has so far seemingly lagged behind China, but it is emerging as a very strong player, alone and in international partnerships, in what is really more a worldwide endeavour than an Asian race.

There are many who question the role of an ambitious space programme in a nation where roughly 450 million people live below the poverty line. Satellites have brought great benefits to India, including connecting remote patients to medical specialists, hooking up tsunami and cyclone warning systems, allowing for better agricultural management and distance education; the Indian Space Research Organisation (ISRO) claims a return of two rupees on every rupee spent on their satellite programme. But there is no doubt that space technology has raced a few paces ahead of the needed social and developmental infrastructure. The bottlenecks are obvious: classrooms equipped with distance-learning satellite linkups are often empty; the ground-based computer linkups required to make telemedicine a reliable option can often collapse. And a Moon mission, unlike the satellite programme, promises no direct good for India's poor.

Yet India's space programme has long benefited from political commitment across party lines. Funding has not been a constraint either. The 2007–08 budget allocates 500 million rupees (US$10 million) for the ISRO to develop a project report for its manned space initiatives.

M. LAKSHMAN/AP

S. BHARADWAJ/DINODIA PHOTO LIBRARY

Vikram Sarabhai saw the utility of space research for India.

How did India end up here? The story begins with the charismatic, trail-blazing father of the Indian space programme — Vikram Ambalal Sarabhai.

Founding father

Sarabhai, born into an illustrious business family in Ahmedabad, Gujarat, was 28 when the country became independent of British rule in 1947. A physicist who had researched cosmic rays with Nobel laureate C. V. Raman and trained at Cambridge, Sarabhai pushed for space-related development in India right from the start. This made all the difference between excellence and mediocrity in a newly independent country grappling with developmental issues of all kinds. In 1947 he founded the Physical Research Laboratory (PRL) in Ahmedabad, using charitable trusts owned by his family and friends to create a centre for research in space and allied sciences. Sarabhai was tireless, and combined a heady mix of creativity and knowledge in both the fine arts and science; in addition to the PRL, he founded the top business school — the Indian Institute of Management Ahmedabad — and helped to set up the National Institute of Design. What made him famous, though, was his vision to see the potential of a space programme for India.

Initially, the PRL focused on researching cosmic rays and the atmosphere. But as space technologies progressed, Sarabhai had the foresight to see how cutting-edge science — particularly communications satellites — could make a real difference to people's lives.

In the late 1950s, the world watched as the Russians sent Sputnik-1 into space, heralding the beginning of a new era and the start of the first space race between the Soviet Union and the United States. India's first prime minister, Jawaharlal Nehru, already had an abiding belief that scientific advancement could solve several of India's problems, and the satellite launch set the country's portals of power buzzing. Nehru asked Homi Jehangir Bhabha, known today as the father of India's atomic-energy programme, to help him craft a scientific policy resolution. Passed in Parliament in early 1958 it stated: "The gap between the advanced and backward countries has widened more and more. It is only by adopting the most vigorous measures and by putting forward our utmost effort into

the development of science that we can bridge the gap." And so, India became dedicated to matching the world's greatest scientific developments.

As Russia's Yuri Gagarin became the first man in space in 1961, India got down to working out the nuts and bolts of its own space programme. In 1962, Sarabhai took charge of the Indian National Committee for Space Research, under the overall guidance of Bhabha. At Thumba, a village in the deep south of Kerala, work began on the Thumba Equatorial Rocket Launching Station that saw its first launch — of a US rocket — in 1963. The aerospace engineer A. P. J. Abdul Kalam, one of the country's original missile men, and later president of India, was part of a small knot of space dreamers who sweated and slogged with Sarabhai in the early sixties — often out of shacks, tin-roof sheds and even an old village church. He still speaks of Sarabhai as one of India's most extraordinary leaders.

Within a few years, the Indian space programme began to focus on the capability to build as well as launch satellites. Sarabhai founded the Space Science and Technology Centre in Thiruvananthapuram, Kerala (now called the Vikram Sarabhai Space Centre) to pursue this work, which culminated in the country's first successful launch into orbit of an Indian satellite in 1980. As well as a great leader, Sarabhai was also a great maker of teams. His work laid a firm foundation for future ISRO leaders to build on.

The Moon mission stands apart from the application-focused satellite programme. Its inception can be traced to Krishnaswamy Kasturirangan, member of parliament and head of the ISRO from 1994 to 2003. Kasturirangan's geniality and easy demeanour is a perfect foil to his sharp and intense vision as a scientist, which gave early shape to the idea of India going to the Moon in 1999. After a series of stakeholder meetings and consultations between 1999 and 2003, a group of 100 eminent scientists officially recommended that India set its sights on the Moon. Their stated goals were to expand human knowledge, and to challenge India to go beyond geostationary orbit, thereby potentially attracting young talent to the space sciences and into the country's space programme.

A new space race is emerging globally, and India clearly wants to be part of that. China and nations within the European Space Agency (ESA) are awaiting approvals for manned missions to the Moon. South Korea is building its first space centre. The Chinese Shenzhou VII, the country's third manned mission into space, took off in September, following its first Moon orbiter in 2007. The United States has cranked up its Moon mission machinery after three decades, and the Russians are developing new generation spacecraft intended for manned missions to the Moon and Mars. Mylswamy Annadurai, project director for Chandrayaan-1, has written: "Decades from now, human colonies on the Moon could become a reality. India should also be in the forefront of this challenging and exciting endeavour."

Grand ambitions

Kasturirangan calls Chandrayaan-1 a "forerunner of more ambitious planetary missions in the years to come". On its heels is Chandrayaan-2, an Indo-Russian robotic rover planned to be deployed from a lunar orbiter in 2011–12, to probe the Moon's surface for geological data and look for helium-3, a potential fuel for futuristic fusion reactors. A manned mission to the Moon is part of the long-term plan, along with an unmanned mission to Mars slated for 2015, and a future mission to an asteroid. The Aditya mission is aimed at the Sun for 2014.

"A new space race is emerging globally, and India clearly wants to be part of that."

India knows that the name of the game inspace science is international collaboration, and there are signs that others will co-operate with the country. In September 2007, NASA administrator Michael Griffin told Pallava Bagla, my co-author on our book *Destination Moon*: "India and the United States are not racing to go the Moon. We can hope to go together."

But as rocket technologies are closely-guarded national secrets, the country sees the need for its own low-cost access to space. As G. Madhavan Nair, chairman of the ISRO and the man who delivered the Moon mission, says: "Twenty years from now, when space travel is likely to become mundane like airline travel today, we don't want to be buying travel tickets on other people's space vehicles." Chandrayaan-1 carries 11 payloads: five Indian, two from NASA, three from the ESA and one from Bulgaria. But it will be launched with the all-Indian technology of the Polar Satellite Launch Vehicle. In the country's other main launch technology — the Geosynchronous Satellite Launch Vehicle — a vital section originally provided by Russia has now been re-made in India, aiming to eliminate reliance on foreign countries for launch capabilities.

More than four decades ago, Sarabhai remarked: "There are some who question the relevance of space activities in a developing nation. To us, there is no ambiguity of purpose." He saw the great benefits to be gained from satellites aiding in the development of the country, as do the heads of the space programme today. But Sarabhai went on: "We do not have the fantasy of competing with the economically advanced nations in the exploration of the Moon or the planets or manned space-flight."

Whether there still is no such fantasy is difficult to tell. India is now aiming at these loftier goals, but often in co-operation, rather than competition. Sarabhai died suddenly in his sleep at the age of 52, succumbing, some say, to the rigours of his career. But reading his lectures and writings, it would be fair to assume that he would be happy with the direction of India's current space programme. His main goal was for his country to be at the cutting edge of space science and technology. Moon missions are now part of that. No wonder Kasturirangan says: "It is not a question of whether we can afford to go to the Moon. It is whether we can afford to ignore it." ■

A. SOLANKI/AP

Millions struggle to survive, yet India invests heavily in its space programme.

Subhadra Menon is an award-winning writer based in New Delhi, and co-author of the book *Destination Moon*.
e-mail: subhadra.menon@gmail.com

For further reading see http://tinyurl.com/4yljl3.

Watch www.nature.com/news for updates on the Moon launch.

ESSAY

D. PARKINS

Santa Fe 1986: Human genome baby-steps

The 1980s saw plenty of discussion on sequencing the human genome. But, according to **Charles DeLisi**, one conference was crucial for converting an idea to reality.

It was summertime. I had left my job as a senior investigator at the US National Institutes of Health (NIH) to take up a new role as head of health and environmental research programmes at the Department of Energy (DoE). This was not an obvious platform from which to launch an effort to sequence the human genome. In fact, organizing such a project was far from my mind. I did, however, feel strongly that we needed a deeper understanding of the impact of human activities, especially energy strategies, on the global environment.

Central to our mission at the DoE was to understand the effect on human health and the environment from energy by-products, such as fuel emissions. I was especially interested in how modern genetics could be brought to bear on these issues. During my decade at the NIH we would occasionally discuss the nature of resistance and susceptibility to disease among humans. At the DoE, the same thought seemed to be emerging in an entirely different context — how can we genetically characterize variations in the susceptibility of individuals to low levels of energy by-products?

In October 1985, two months after I arrived at the DoE, David Smith, a senior member of the health and environmental research staff, handed me a copy of a report. It was called *Technologies for Detecting Heritable Mutations in Human Beings* and had been written by the now defunct Office of Technology Assessment, whose function was to provide science advice to members of the US Congress.

Not surprisingly, the report had been informed by an earlier meeting on this very topic. That meeting had been organized by DoE staff — David Smith himself and Mortimer Mendelsohn, then head of health and environmental research programmes at the Lawrence Livermore National Laboratory in California. It was clear from this report that real progress towards meeting DoE goals would require bolstering our genetic programmes, and the report hinted at how this might be done.

The idea was to sequence the human genome. It was known at the time that, on average, the genetic difference between two individuals was approximately one base per thousand. So if we were able to sequence one genome, this could act as a reference point for information on genetic differences. I immediately called Mendelsohn, who headed our advisory board at the DoE, and asked him what he thought.

He told me about another meeting organized five months earlier by Robert Sinsheimer at the University of California in Santa Cruz. Sinsheimer, a former biologist at the California Institute of Technology in Pasadena and chancellor of the University of California, wanted to explore the feasibility of sequencing the human genome. I also found out that there had been an earlier conference of scientists in Alta, Utah, to discuss emerging sequencing technologies — although this meeting made no proposal to sequence a full genome.

I then discovered that there was strong interest from leaders in the field, such as Leroy Hood, Walter Gilbert and Charles Cantor, but that the path from idea to implementation was far from clear. Crucially, a human genome project would need substantial funding. Yet the

major funding agencies appeared uninterested, and the idea languished.

Politics and money

Smith and I were eager to sample a broader cross-section of the community, and I asked physician Mark Bitensky, head of life sciences at the Los Alamos National Laboratory in New Mexico, to organize a workshop for leading lights in molecular genetics and allied fields, including the computational sciences. The brief was to assess the costs, value and feasibility of a human genome project, as well as the time needed to complete it. Bitensky was an obvious choice as he had a strong interest in the issues surrounding genetic variation from the point of view of personalized medicine.

The delegates, who gathered in Santa Fe on 3–4 March 1986, included many who had been involved in the earlier Office of Technology Assessment report, as well as several who had attended the Alta summit the previous year. Other prominent geneticists were also present, as were representatives from industry.

Not all were familiar with the history of the idea of sequencing the human genome and, as had been the case at previous meetings, not everyone was comfortable with it. Early on, discussions about costs, organizational architecture and technical obstacles were extremely spirited. There was widespread disagreement over whether it made sense to sequence a genome, more than 80% of which we already knew was non-coding. There was also great concern over the ability to achieve sufficiently low error rates at reasonable cost, and over the US$3-billion price tag. Not to mention the boring and repetitive nature of the project.

There was, however, unanimity about the project's potential value to science (gene regulation, developmental biology, evolution) and to medical applications such as genetic disease and cancer. Interestingly, there is no record of discussion on what is perhaps the greatest beneficiary of the sequencing revolution — infectious disease.

The meeting reached a broad consensus on almost all issues. And although no definitive recommendation was made on how the project would be organized, delegates unanimously recommended forming a steering committee that would help shape a management plan.

The Santa Fe meeting was a success with scientists and also helped to open doors to prospective funding sources in government. Invitations had been sent out to heads of federal agencies asking them to send representatives to the meeting. Only one response came back, and that was an expression of regret. But after the conference, things began to change. The idea of a full genome sequence was now on the national scientific stage, which enabled us to begin the critical task of garnering support from the DoE, the Reagan administration and Congress.

In late April, based on letters I received from participants of the Santa Fe workshop, and on a report prepared by Bitensky, I wrote a memo to Alvin Trivelpiece, the assistant secretary at the DoE to whom I reported. In this I outlined a project that would be divided into three phases: technology development, mapping and sequencing. I also formed a genome advisory committee that included Francis Collins, who went on to head the US Human Genome Project.

Enter the NIH

With Trivelpiece on board, we began to alert Congress and the White House Office of Management and Budget (OMB) to discussions in the scientific community, including the DoE's plans for genome sequencing. The OMB was surprisingly supportive, considering the huge cost and its reputation as a fierce budget-cutting organization. Officials were impressed, both with the unanimity of the Santa Fe workshop on an idea that was unusually ambitious and potentially paradigm altering, and with the fact that this would not be an open-ended science project. It was deliberately pitched in a way that said: this is an engineering/infrastructure-type activity. It has an end-point and well-defined milestones.

I also developed a rapport with the Republican senator from New Mexico, Pete Domenici. Domenici, being from a state that housed two major national laboratories, Sandia and Los Alamos, was accustomed to dealing with abstruse physics projects, and was pleased to have before him a project whose relevance could easily be explained to his constituency. As a member of the Senate's budget committee and a ranking member of the powerful appropriations subcommittee on energy and water development, we needed Domenici to obtain the support of Congress and Administration to move the project forwards.

> "Discussions about costs and technical obstacles were extremely spirited."

With support from the Secretary of Energy and the OMB, a $13-million line item initiating the genome project appeared in President Reagan's budget submission to Congress in January 1987. It subsequently passed both Houses, and 1988 saw the first official expenditures on the Human Genome Project.

As the DoE moved forwards during 1986, word of an unprecedented initiative was spreading. At a June 1986 Cold Spring Harbor meeting on the 'Molecular biology of *Homo sapiens*', sequencing the human genome became the topic for an impromptu discussion.

Unlike at Santa Fe, Cold Spring Harbor heard more voices urging caution. Several participants believed that the project would lead to masses of unevaluated data, or that the computational methods available to us at the time would yield relatively little information. This was part of a more general concern that the technology of the day was not appropriate for the complexity of the task.

Other researchers feared that the project would be subject to political interference, as was sometimes seen with NASA, and regarded the DoE as the wrong agency to manage it. For its part, the NIH was concerned legitimately that a large project, spread out over more than a decade, would shift substantial sums of money away from worthwhile investigator-initiated proposals. Nevertheless, James Watson was among those who felt that NIH involvement was crucial. In the summer of 1986, he successfully persuaded Congress to include a human genome allocation as part of the agency's 1988 budget.

I left the DoE in the summer of 1987, feeling naively certain that the project was in safe harbour, and that a complete sequence would be ready by our target date of 2001. We got the date right, but for the wrong reasons. The 2001 date was based on an assumption that the economy would be relatively normal. In fact, the mid-1990s was an incredibly vibrant period economically and stimulated investments from venture capitalists, some of whom made possible the formation of Celera Genomics, the company led by Craig Venter that was at the head of a private-sector sequencing effort. Without the ensuing public versus private competition, it is unlikely that the complete sequence would have been ready by 2001 because the target date was reset to 2006 after I left Washington.

As in all complex human ventures, the Human Genome Project had its share of stresses, squabbles and power plays between different agencies. In the end, everyone that mattered pulled together to create what I regard as a phase transition in national science policy. It is a monumental tribute to the biomedical research enterprise in the United States — not only to the scientific ingenuity that brought it successfully to completion, but to a culture of versatility and adaptability. ■

Charles DeLisi is Metcalf professor of science and engineering at Boston University.
e-mail: delisi@bu.edu

For more Meetings that Changed the World, see www/nature.com/nature/focus/meetings.

nature

NEWS & VIEWS

STRUCTURAL BIOLOGY

Clamour for a kiss

Anastassios Economou

Precisely how proteins snake their way through channels in cell membranes is unclear. Complexes between the SecY channel and its motor protein, and the use of a 'molecular endoscope', provide fascinating clues.

To cross cellular membranes, secretory proteins use specialized channels such as that known as SecY. When ribosomes start synthesizing secretory proteins, the ribosome–protein complex is hijacked to the mouth of SecY. There the nascent 'pre-protein' chains, which carry specific recognition tags known as signal peptides, are threaded through the endoplasmic reticulum — a membranous mesh inside the cells of eukaryotic organisms — or through the plasma membrane that surrounds prokaryotic cells such as bacteria. In bacteria, a 'push' motor called SecA can also mediate ribosome-independent protein secretion through SecY channels, using the cell's energy currency ATP in the process. The structure of such cellular machines dictates their function and so can be highly informative. But despite having known for some time[1,2] what the separated parts — SecY and SecA — look like, we are no wiser about what they look like when united. In this issue, three studies[3–5] provide eagerly awaited snapshots of how the SecY channel and the SecA motor come together in a 'kissing' complex, and how they might move to mediate protein translocation.

SecY channels come in pairs in which each has its own specific functions — one half seems to carry out the actual translocation of proteins, the other holds the SecA motor in place[6]. Near the cytoplasmic entrance of both channels are several docking sites for ribosomes and motors, and at their exits a plug structure[1,7]. Motors also come in pairs and each copy comprises four domains: two make up the 'engine' and sandwich ATP molecules between them[2]; the other two form the 'business end' for translocation and can be thought of as 'hands' (Fig. 1a).

Zimmer *et al.*[3] (page 936) share the first glimpse of a single bacterial protein-conducting channel in complex with a single SecA motor (Fig. 1b) — a technical feat because detergent extraction of proteins from membranes often dissociates them from their binding partners. The channel seems to have undergone conformational changes compared with its closed state[1], including loosening of its plug. Moreover, one of the motor hands has swivelled around its slender stem in such a way as to delimit a tubular enclave, forming a continuum with the channel. Pre-protein

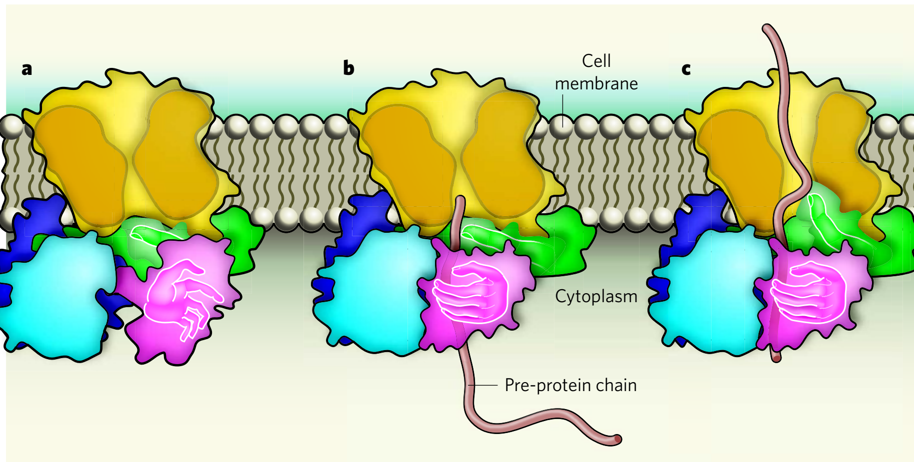


Figure 1 | Protein translocation in bacterial cells. This simplified representation is based on both earlier studies and the new findings[3–5]. In this cut-away view of the membrane, the SecA motor lies flat against the cytoplasmic side of the SecY channel (yellow), and consists of a two-domain ATP-powered engine (light and dark blue) and two 'business-end' domains (green and magenta; depicted as hands). **a**, Initially, the channel pore is sealed by both a constriction halfway through it and a mobile plug domain (not shown) near its exit. The pre-protein-binding domain of the motor (magenta) is in the open state, exposing an elongated corridor that connects to the entrance of the channel. This open state is seen in structures of the isolated motor. **b**, Swivelling this domain around its stem would allow it to embrace a secretory protein chain. At this stage, a finger (green) from the second hand of SecA might be in close contact with the chain. **c**, When ATP (not shown) is present, the engine conformation changes and the finger could move upwards, pushing or dragging the protein chain into the pore. This motion, or other conformational changes, leads to the opening of the pore.

chains interact with this region of the motor[8] and — although not present in this structure[3] — can be imagined to occupy it. For this to occur, the enclave, which is tightly packed and extremely narrow in the present structure, would have to open up. The authors[3] find that a two-helix finger from the motor's other hand dips near the cytoplasmic mouth of the channel, next to where the pre-protein chain would be. This finger is an essential switch that couples ATP expenditure to protein movement[9]. Could it somehow drive the chain through the channel?

To test this possibility, Erlandson *et al.*[4] (page 984) used the pre-protein as an endoscope, trapping it within the channel to study what it 'sees' around it. They constructed mutated derivatives of the pre-protein, the channel and the motor that contained strategically placed cysteine residues. When two copies of this amino acid are juxtaposed within roughly 5 angstroms of each other, even if they are on different proteins, they can form covalent bonds and so reveal proximity. In this case, detergent extraction is unnecessary, and the complex remains fully functional.

These analyses show that, at first — when the motor is idle — specific residues from the pre-protein are in proximity to the 'fingertip'. In a subsequent ATP-driven step, these residues are carried well into the channel (Fig. 1c). On the basis of the chain lengths these authors[4] used, it would be expected that only short segments of the pre-protein — which would have to be stretched — can travel forwards in each cycle. Several repeats of this motion would be required for a complete pre-protein chain of a hundred to hundreds of residues to be pushed across to the other side. Thus a flick of the fingertip could provide a mechanical motion that drags or pulls the trapped pre-protein forwards. Alternatively, the finger could prise the channel

NATURE|Vol 455|16 October 2008

open and/or prevent back-slippage, allowing forward movement of pre-protein segments by Brownian motion. Detailed kinetic assays will be necessary to discriminate between these possibilities.

To visualize the channel–motor complex, Zimmer *et al.*[3] had to immobilize it in a crystal lattice. Consequently, they could capture the structure of only one of its possible states — much like a party snapshot freezes a single move on the dance floor. Synthesis of the full spectrum of channel and/or motor motions will necessitate more snapshots. Tsukazaki *et al.*[5] present one such additional picture on page 988, revealing a different conformational state of the SecY channel.

In these authors' structure, the motor is absent. Instead, they used an antibody to hold a single SecY channel in a particular state. The antibody binds specifically to a motor-binding site and freezes the channel in what is proposed to be a 'pre-open' state. This conformation exposes a cleft with access to the cytoplasm, which Tsukazaki and colleagues hypothesize[5] could attract signal peptides. Whether these antibody-driven effects faithfully mimic those driven by the binding of the motor is unknown. Nevertheless, they reveal the inherent repertoire of potential channel motions.

Like Erlandson and colleagues[4], Tsukazaki *et al.* also used cysteine cross-links to study a functional complex of the channel with the motor. These experiments revealed that a region of the motor undergoes significant conformational changes when it binds to the channel. These interactions probably involve the second copy of the channel, which is absent from Zimmer and colleagues' channel–motor structure. This second copy is thought[1] to act as a docking station for the engine part of the motor.

The implications of these channel–motor 'kiss-and-tell' structures for understanding the mechanics of pre-protein translocation are important. A possible pathway for the pre-protein to follow has now been coarsely charted. The range of possible channel conformations and the motions of the motor's domains can now be predicted and tested. A second copy of the motor is apparently not required for interaction with a single translocating channel. Similarly, the remarkable ATP-driven motions of the motor's hands that accompany protein secretion[10] need not lead to profound membrane penetration of the motor. But to confirm these and other conclusions, experimentally less disruptive isolation of the complexes in a functional form and freezing of other conformational states will be necessary.

Solving structures of membrane proteins is not a trivial pursuit. As we contemplate in awe the reported achievements[3–5], a wish list of future findings comes to mind: higher resolution structures of the twin channel–motor complex, accompanied by a trapped pre-protein chain; understanding how the motor recognizes pre-proteins and how it converts the chemical energy of ATP into mechanical work; and determining the dynamics of this astonishing cellular nanomachine. ■

Anastassios Economou is at the Institute of Molecular Biology and Biotechnology, Foundation of Research and Technology-Hellas, and in the Department of Biology, University of Crete, PO Box 1385, Iraklio, Crete, Greece.
e-mail: aeconomo@imbb.forth.gr

1. van den Berg, B. *et al. Nature* **427,** 36–44 (2004).
2. Hunt, J. F. *et al. Science* **297,** 2018–2026 (2002).
3. Zimmer, J., Nam, Y. & Rapoport, T. A. *Nature* **455,** 936–943 (2008).
4. Erlandson, K. J. *et al. Nature* **455,** 984–987 (2008).
5. Tsukazaki, T. *et al. Nature* **455,** 988–991 (2008).
6. Osborne, A. R. & Rapoport, T. A. *Cell* **129,** 97–110 (2007).
7. Mori, H. & Ito, K. *Proc. Natl Acad. Sci. USA* **103,** 16159–16164 (2006).
8. Cooper, D. B. *et al. J. Mol. Biol.* **382,** 74–87 (2008).
9. Karamanou, S. *et al. Mol. Microbiol.* **34,** 1133–1145 (1999).
10. Economou, A. & Wickner, W. *Cell* **78,** 835–843 (1994).

CONDENSED-MATTER PHYSICS

Surviving the transition

Kristian Helmerson

Observations of the birth of a superfluid have uncovered details of the microphysics of phase transitions. Whether these results can be used to model such transitions in the early Universe is an open question.

The spontaneous formation of topological defects — stable configurations of matter — is thought to accompany most continuous, non-equilibrium phase transitions in condensed-matter physics. But the study of the microscopic dynamics underlying the formation of such defects is a challenge, especially in superfluid transitions. On page 948 of this issue, Weiler *et al.*[1] report observations of topological defect formation in a particularly helpful superfluid phase transition — the Bose–Einstein condensate (BEC) transition— in a trapped gas of rubidium atoms. These observations offer unique insights into the microscopic physics of phase transitions.

Approximately 10^{-35} seconds after the Big Bang, the Universe underwent a phase transition resulting from the cooling associated with its rapid expansion. We generally think of a phase transition driven by cooling as one that takes a system from a disordered state to an ordered one — for example, water turning into ice. Yet the Universe is not an ordered state, but is filled with galaxies, stars and other celestial bodies. A way out of this dilemma was proposed by Thomas Kibble[2,3] in 1976. He argued that, as the Universe cooled and approached the phase transition, large fluctuations in the vacuum in the form of topological defects, such as cosmic strings, were formed. These defects, which persisted even after the transition, eventually led to the formation of larger structures in the Universe.

The key to the survival of these topological defects is that they are formed locally on a timescale that is shorter than the time it would take light to propagate from the location of one defect to another; hence the defects would not be able to know about one another. In the parlance of physics, this means that the formation of the defects is not causally connected. Using similar reasoning, Wojciech Zurek suggested[4,5] that in a condensed-matter system, a second-order phase transition — in which the transition between the old and the new phase is continuous — could result in defect formation if the transition is crossed rapidly enough.

Zurek went on to propose a specific experiment involving the phase transition of liquid helium ^{4}He, at a temperature of 2.17 Kelvin, from a normal fluid to a superfluid; that is, a fluid with no viscosity. If the helium could be cooled rapidly enough, then topological defects in the form of quantized vortices would be produced. In this case, however, the timescale for the causal connection of the defects is not determined by the speed of light but by the speed of sound in the bulk superfluid.

Since Zurek's suggestion, the mode of formation of topological defects, which is known as the Kibble–Zurek mechanism, has been observed in liquid crystals, arrays of superconducting Josephson junctions, nonlinear optical systems, fluid convection systems and superfluid ^{3}He. But, surprisingly, it has not been observed in ^{4}He, the quintessential superfluid and the system referred to in Zurek's original proposal. Weiler *et al.*[1] now report observing such defects in a BEC transition of a trapped atomic gas.

In their experiments, Weiler *et al.* evaporatively cooled a gas of weakly interacting rubidium (^{87}Rb) atoms confined in a magnetic trap. Evaporative cooling normally occurs slowly, such that the system being cooled is close to reaching thermal equilibrium. This is similar to what occurs when the hottest (highest-energy) molecules evaporate from a cup of coffee and the rest of the coffee 're-thermalizes' — that is, comes back to thermal equilibrium — at a lower temperature. The system can, however, be thrown out of equilibrium by rapidly modifying the energy distribution. Weiler *et al.* achieved this by using radio-frequency transitions that

selectively remove a substantial fraction of the highest-energy atoms from the magnetic trap. This technique was first used[6] to study the formation dynamics of a BEC.

Weiler and colleagues found a regime in which the gas of rubidium atoms can rethermalize locally on a timescale faster than it would take for atoms on one side of the magnetic trap to let atoms on the other side know what is going on. As a result, a local mini-BEC could form in one place in the trap while another mini-BEC formed elsewhere. Because these mini-BECs (typically referred to as quasi-condensates) do not know about each other, they will each have a quantum-mechanical phase that is independent of one another. When three such mini-BECs with independent phases merge, they can form a vortex, as an earlier experiment[7] showed.

A more formal description of the process of vortex formation would characterize the size of the quasi-condensates by the coherence length, ξ (the spatial extent over which the phase is uniform). If the phase transition is approached while the system is close to thermal equilibrium, ξ diverges at the critical point of the phase transition and there is no length scale for quasi-condensate formation. Thus, no topological defects are expected. But if the critical point is crossed rapidly enough, then the coherence length just before crossing the transition is frozen in, and quasi-condensates of coherence length ξ can form, which, for a big enough system, would lead to the formation of topological defects.

Although the experiments of Weiler *et al.* support the mechanism of Kibble and Zurek, the finite size of the system limits the observation of universal scaling laws, such as the density of defect formation, which should increase as $1/\xi^2$. However, given that non-equilibrium phase transitions have important consequences, as exemplified by the Kibble–Zurek mechanism, these experiments are a welcome addition. Such experiments[1,8] show that trapped atomic gases provide a system in which the microscopic dynamics underlying the important mechanisms in a non-equilibrium phase transition can be probed, understood and controlled. Whether they can be used to model the early evolution of the Universe remains to be seen, but they look promising for providing new insights into these, traditionally difficult to study, phase transitions. ■

Kristian Helmerson is at the National Institute of Standards and Technology, Gaithersburg, Maryland 20899-8424, USA.
e-mail: kristian.helmerson@nist.gov

1. Weiler, C. N. *et al. Nature* **455,** 948–951 (2008).
2. Kibble, T. W. B. *J. Phys. A* **9,** 1387–1398 (1976).
3. Kibble, T. W. B. *Phys. Rep.* **67,** 183–199 (1980).
4. Zurek, W. H. *Nature* **317,** 505–508 (1985).
5. Zurek, W. H. *Phys. Rep.* **267,** 177–221 (1996).
6. Miesner, H.-J. *et al. Science* **279,** 1005–1007 (1998).
7. Scherer, D. R., Weiler, C. N., Neely, T. W. & Anderson, B. P. *Phys. Rev. Lett.* **98,** 110402 (2007).
8. Sadler, L. E., Higbie, J. M., Leslie, S. R., Vengalattore, M. & Stamper-Kurn, D. M. *Nature* **433,** 312–315 (2006).

EARTH SCIENCE

Deducing a reducing mantle

William F. McDonough

Increasingly sophisticated techniques are being used to persuade ancient rocks to yield information about conditions on and in the early Earth — for instance, about the oxidation state of the mantle.

They say that a picture is worth a thousand words. In science, one datum point can be worth a thousand models. On page 960 of this issue, Berry *et al.*[1] report the first results obtained by applying a new approach to estimating the oxidation–reduction (redox) condition of Earth's upper mantle during the Archaean, some 2,700 million years ago. Their study is based on the ratio of different oxidized states of iron (Fe^{3+} to Fe^{2+}) in inclusions of ancient komatiite rock, trapped in crystals preserved in ancient lava flows in Zimbabwe. The crystals should have protected the inclusions from subsequent alteration, which should thus reflect the native lava state.

Berry and colleagues' findings are groundbreaking, both for the experimental method they used (X-ray absorption near-edge structure spectroscopy, which shows that we can interrogate minute and precious fragments of the ancient Earth), and for what the findings reveal about the early Earth. They will embolden others to analyse rare recorders of ancient and remote redox conditions, such as ancient zircons and other minerals, early Solar System condensates in meteorites, and inclusions in diamonds. And they have implications for several aspects of our understanding of Earth's history: the secular evolution of mantle oxidation state; oxidation of the atmosphere; melting conditions for the formation of the ancient lavas; and the nature of the tectonic environment in which these lavas formed.

When, around 2,700 million years ago, the lavas erupted on the Earth's surface, the atmosphere was low in oxygen, and the circumstances that first produced an oxygenated atmosphere (the 'great oxidation event') were still 500 million years off. Ever since the first conjectures about the great oxygenation event[2,3], authors have speculated as to whether the mantle was the source of the oxygen. But the redox conditions of the mantle today do not necessarily reflect those in the early Earth because of the degassing, magmatism, cycling of tectonic plates and mantle convection that have occurred during the past 2,700 million years.

The komatiitic melt inclusions studied by Berry *et al.*[1] have low concentrations of Fe^{3+} (that is, a high degree of reduction and low oxidation), which the authors interpret as reflecting the redox state of the host lavas. That state is comparable to the redox state of the mantle source from which the continents are formed — basalt rocks produced at mid-ocean ridges, where tectonic plates are diverging. These rocks have the most highly reduced condition recorded for major outpourings of basaltic magmas, as compared with other tectonic settings (for example convergent boundaries, where one plate is being thrust beneath another, or where volcanic activity is occurring within a plate).

Today, basalts from convergent, Andean-type margins, which include those from the Pacific's ring of fire, are the most highly oxidized, and this is thought to be due to the addition of water transported along with the down-going, subducted plate to the site of magma generation. Basalts from intraplate settings, such as those beneath Hawaii, have intermediate water content and a transitional oxidation state between those of basalts from Andean margins and from mid-ocean ridges. Berry and colleagues show that their komatiite melt inclusions are not significantly water-bearing. Thus, the low Fe^{3+} and low water content of the inclusions are consistent with komatiite genesis through large-scale melting of the mantle under relatively anhydrous, high-temperature conditions[4,5].

Because these ancient, Archaean melt fragments record such a reduced oxidation state for iron, one comparable to that of the dominant volume of mantle today, the implication is that the mantle is unlikely to have provided the atmosphere's oxygen in the past. It is not impossible that it did; but such an outcome would have required exceptional circumstances.

The early mantle was hotter and more active than it is today, as evidenced by the presence of komatiites (such high-temperature melts are limited to Earth's first 2,500 million years or so); the higher content of heat-producing radioactive elements; and the greater impact flux of asteroids in the first thousand million years of Solar System history. Thus, mantle convection was more vigorous in the early Earth, which probably led to a more homogeneous mantle, one less capable of storing unprocessed, more-oxidized material. Given this hot, active, early Earth, the findings of Berry *et al.* point to a mantle that became more oxidized, not more reduced, with time. This in turn could indicate that the process of recycling oceanic plates back into the mantle perhaps contributed to the progressive oxidation of part of the mantle through time.

The tectonic context for komatiite genesis continues to be a puzzle. Berry and colleagues' results, coupled with an earlier trace-element study on similar melt inclusions from these lavas[6], reveal that these high-temperature melts are more akin to modern lavas from a divergent

margin or from an intraplate setting than they are to lavas from convergent margins. In contrast, komatiitic lavas from the ancient Superior Province of Canada are stratigraphically associated with basalts and other lavas from a convergent margin. Moreover, the discovery of komatiitic deposits of explosively ejected pyroclastic ash, interlayered with lava flows that occurred on land or in shallow water[7], document the eruption of these rock suites in terrestrial and near-terrestrial settings. Thus, it is possible that komatiite genesis occurred at divergent margins, but not at margins associated with deep, mid-ocean-ridge environments.

More work is needed if we are to find and analyse melt inclusions from other komatiites and ancient lavas, with the aim of building a comparative perspective on the secular evolution of the oxidation state of lavas from various tectonic settings. The analytical methodology applied by Berry *et al.*[1] is now available for studying minute geological time-capsules, and so opens the door for many other studies. There are plenty of old hypotheses to be tested with this new tool. ■

William F. McDonough is in the Department of Geology, University of Maryland, College Park, Maryland 20742-4211, USA.
e-mail: mcdonoug@geol.umd.edu

1. Berry, A. J., Danyushevsky, L. V., O'Neill, H. St C., Newville, M. & Sutton, S. R. *Nature* **455,** 960–963 (2008).
2. Holland, H. D. in *Early Life on Earth* (ed. Bengtson, S.) 237–244 (Columbia Univ. Press, 1994).
3. Farquhar, J. *et al. Science* **289,** 756–758 (2000).
4. Arndt, N. *et al. Geology* **26,** 739–742 (1998).
5. Herzberg, C. *J. Geophys. Res.* **97,** 4521–4540 (1992).
6. McDonough, W. F. & Ireland, T. R. *Nature* **365,** 432–434 (1993).
7. Stiegler, M. T., Lowe, D. R. & Byerly, G. R. *Geology* **36,** 779–782 (2008).

CANCER

A ringleader identified

Charis Eng

The childhood cancer neuroblastoma can either run in families or occur sporadically. Several studies find that the gene *ALK* is a chief offender in this disease, because its germline mutations mediate both forms.

Neuroblastoma is the most common childhood cancer diagnosed before the age of one, and accounts for some 15% of all cancer deaths in children. Some patients inherit a genetic predisposition to neuroblastoma due to mutations in the germ line (cells that can give rise to all other cells in an organism); in others, the cancer is sporadic, that is, it may be caused by both germline and somatic (found only in tumour cells) mutations. In this issue, four studies[1–4] implicate various mutations in the gene *ALK* in both familial and apparently sporadic cases of neuroblastoma.

Neuroblastoma tumours are derived from embryonic cells that form the neural crest and that eventually give rise to the peripheral nervous system, among other organ systems. The first gene to be linked to predisposition to neuroblastoma was *PHOX2B* — a master regulator of the normal development of the autonomic nervous system[5,6]. But the *PHOX2B* germline mutations identified occur only in a small subset of familial neuroblastoma cases, and somatic mutations in this gene have not been reported. Furthermore, germline *PHOX2B* mutations result in syndromic neuroblastoma — meaning that the cancer associates with other, very specific clinical features. So it seemed that germline mutations in other, unidentified genes were responsible for most cases of familial neuroblastoma, and the search went on.

After a few red herrings, a clue came from an unexpected quarter: the short arm of chromosome 2 was found to be somatically amplified in a subset of sporadic neuroblastoma cases. This chromosomal region contains both the *MYCN* gene, amplification of which is well established for sporadic cases of neuroblastoma, and *ALK*.

The protein product of *ALK* is a tyrosine kinase, an enzyme that regulates the activity of other proteins through phosphorylation. Abnormal activity of the ALK protein has been implicated in the amplification and translocation of genomic sequences in non-Hodgkin's lymphoma and various solid tumours, including non-small-cell lung cancers. It now emerges that *ALK* is also a major non-syndromic neuroblastoma-predisposition gene.

Mossé and colleagues[1] (page 930), Janoueix-Lerosey *et al.*[2] (page 967) and Chen *et al.*[3] (page 971) together identify four different germline *ALK* mutations in 10 of 16 neuroblastoma pedigrees they studied. Of these, the frequently occurring mutations do not seem to represent founder mutations — which occur in a common ancestor and perpetrate through subsequent generations — but rather mutational 'hotspots'. Hotspots are characteristic of gain-of-function mutations in cancer-predisposing oncogenes[7] that lead to increased activity of the oncoprotein.

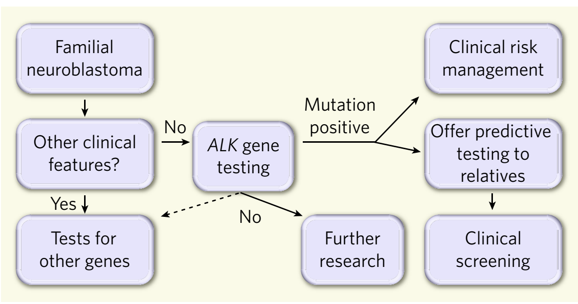
Bearing in mind that the number of unrelated families with neuroblastoma studied by these authors[1–3] is relatively small, the weighted mean proportion of individuals carrying the germline mutations who actually develop neuroblastoma (penetrance) seems to be around 57% across all families. One of the germline mutations (substitution of alanine in place of glycine at amino-acid residue 1128) detected in a single large family[1] seemed to have a lower penetrance of about 40% compared with the other three mutations, whose weighted-average penetrance was 61%. These penetrance percentages are relatively low compared with those of other established cancer-predisposition genes such as *RET* and *PTEN* (ref. 7).

Box 1 | Pre-emptive measures to manage neuroblastoma

With the identification of *ALK* as a main neuroblastoma-predisposition gene[1–4], individuals whose families harbour specific *ALK* mutations but who have no clinical features suggestive of neural-crest defects should be offered tests for mutations in this gene as part of genetic counselling. This has certainly proved to be a useful approach in the case of *RET* mutations, because analysis of certain coding sequences within this gene allows sensitive and cost-effective genetic testing for predisposition to a form of cancer called multiple endocrine neoplasia type 2.

Like ALK, RET is a tyrosine kinase expressed in cells derived from the neural crest. Because only a limited number of hotspot gain-of-function *RET* mutations are germane in the pathogenesis of multiple endocrine neoplasia type 2 (ref. 7), by inference it is unlikely that future studies will identify several other mutation hotspots associated with neuroblastoma. So once a family-specific *ALK* mutation is identified, all as-yet-unaffected first-degree relatives of the patient can be offered single-genetic-site testing for that mutation, shortly after birth, to determine their risk of developing neuroblastoma compared with that of the general population.

Those with the mutation, and so at risk, can then be offered organ-specific clinical screening so that tumours can be detected at an earlier, more treatable, stage. **C.E.**



George *et al.*[4] (page 975) identify one of the same four germline mutations[1–3] in 2 of 96 patients with apparently sporadic neuroblastoma (as other members of the families to which these two individuals belonged were not examined). Moreover, these studies[1–4] report that 59 of 617 (9.6%) sporadic neuroblastoma cases they investigated have somatic single-nucleotide mutations in *ALK*. Such somatic mutations were associated with more aggressive tumours and lethal cases of this cancer.

These studies[1–4] also found that, in contrast to normal ALK, the mutated enzyme was variably phosphorylated and had increased kinase activity in a mutation-specific way. Similarly, downstream targets of ALK were activated by the mutations in this enzyme in a mutation-dependent manner. Could the increased activity of ALK in neuroblastoma be inhibited for treatment purposes? For example, a leucine for arginine substitution at position 1174 of ALK results in phosphorylation of the STAT3 and AKT proteins, whereas the replacement of arginine by glutamine at position 1275 was associated with phosphorylation of AKT and the ERK1/2 protein. Although these data hint that ALK inhibition[3,4] could be a viable strategy for the treatment or prevention of neuroblastoma, they also indicate that multi-agent targeted therapy as well as nonspecific kinase inhibitors make sense, given that multiple signalling pahways consisting of several kinases are involved.

Mossé *et al.*[1] and Janoueix-Lerosey *et al.*[2] find that although members of 6 of the 16 families they examined do not harbour mutations in *ALK* and *PHOX2B*, neuroblastoma runs in these pedigrees. At least three possibilities, which are not mutually exclusive, could explain these observations. First, mutations in other genes, such as *MYCN*, might be involved. Second, because ALK-mediated neuroblastoma involves increased activity of this protein, germline mutations in promoter sequences that favour *ALK* expression are possible.

Third, large genomic deletions and rearrangements could occur in the germ line. Somatic translocations involving *ALK* have been reported[8], and so germline rearrangements in the sequence of this gene are also plausible. *ALK* deletions associated with neuroblastoma seem counter-intuitive, however, because it is increased ALK activity — rather than its absence — that seem to lead to this cancer. But germline deletions in regions containing repressors of *ALK* expression, as well as partial deletions creating ALK-like proteins, new proteins distinct from ALK, or continuously active kinases, are among plausible mechanisms.

Chen *et al.*[3] and Mossé and colleagues[1] provide evidence that somatic *ALK* mutations associate with aggressive forms of neuroblastoma in sporadic cases. Several questions arise from these observations. For example, will somatic mutations make *ALK*-associated familial neuroblastoma more aggressive? If the answer is yes, should adjuvant therapy — perhaps ALK inhibitors — accompany surgical removal of the tumour? Could it be that ALK inhibition is toxic rather than beneficial in cancer cases caused by germline mutations? After all, every single cell in the body will carry the mutation, albeit with differential expression in different tissues.

The past two years have seen an explosion in genome-wide association studies, which have shown that variations of certain genes with low penetrance account for a small subset of various common cancers. At present, these data are not associated with much clinical context, and so cannot meaningfully contribute to genetic counselling and cancer management. In this era of genomic medicine, the long-awaited discovery of a major non-syndromic neuroblastoma gene[1–4] is indeed a welcome advance for taking pre-emptive measures (Box 1). ■

Charis Eng is at the Genomic Medicine Institute, Cleveland Clinic, 9500 Euclid Avenue, Cleveland, Ohio 44195, USA.
e-mail: engc@ccf.org

1. Mossé, Y. P. *et al. Nature* **455,** 930–935 (2008).
2. Janoueix-Lerosey, I. *et al. Nature* **455,** 967–970 (2008).
3. Chen, Y. *et al. Nature* **455,** 971–974 (2008).
4. George, R. E. *et al. Nature* **455,** 975–978 (2008).
5. Amiel, J. *et al. Nature Genet.* **33,** 459–461 (2003).
6. Mossé, Y. P. *et al. Am. J. Hum. Genet.* **75,** 727–730 (2004).
7. Zbuk, K. M. & Eng, C. *Nature Rev. Cancer* **7,** 35–45 (2007).
8. Soda, M. *et al. Nature* **448,** 561–566 (2007).

PHILOSOPHY OF SCIENCE

Theories of almost everything

P.-M. Binder

A provocative contribution to the logic of science extends the theorems of Kurt Gödel and Alan Turing, and bears on thinking about prediction, the standard model of particles, and quantum gravity.

Since 1620, when Francis Bacon's *Novum Organum* set out the basic guidelines, the task of science has been to condense multiple observations into brief, general descriptions of natural phenomena. This process, called induction, helps us understand and predict the world around us. Enquiries into the limits of science[1] have involved asking questions such as "Can we know everything about the natural world?", which have so far gone unanswered. Writing in *Physica D*, David Wolpert[2] has made headway in this direction by demonstrating that the entire physical Universe cannot be fully understood by any single inference system that exists within it.

Various major scientific developments of the twentieth century have placed limits on different facets of knowledge. These include the measurement process (quantum mechanics, through Heisenberg's uncertainty principle); the transmission of information (relativity, through the constancy of the speed of light); the ability to predict the future from less-than-perfect measurements in the present (chaos theory, through sensitive dependence on initial conditions); and the efficient prediction[3] of certain natural phenomena before they unfold (complex systems theory, through intractability).

Wolpert's work follows another path, that of Kurt Gödel and Alan Turing's theorems of incompleteness developed in mathematics and computation (Box 1), and he extends them to address the logic of science. He introduces the idea of inference machines — physical devices that may or may not involve human input — that can measure data and perform computations, and that model how we come to understand and predict nature. He develops a formal description of all such inference machines in terms of two functions: one stipulates the initial state of a machine (the set-up function) and the other (the conclusion function) describes the observations, recollections or predictions it makes — in other words, a 'theory'.

In proving his theorems, Wolpert defines U as the space of all world-lines (sequences of events) in the Universe that are consistent with the laws of physics. He then defines strong inference as the ability of one machine to predict the total conclusion function of another machine for all possible set-ups. Finally, he uses 'Cantor diagonalization' (Box 1) to prove, among others, the following two statements:

(1) Let C_1 be any strong inference machine for U. There is another machine, C_2, that cannot be strongly inferred by C_1.

(2) No two strong inference machines can be strongly inferred from each other.

The first of these statements posits that there is a portion of 'knowledge space' (that inferable by C_2) that is not available to any C_1 machine. The second is a statement about the non-equivalence of inference machines; it implies that, at most, only one machine at one

instant in time can infer all others. The two statements together imply that, at best, there can be only a 'theory of almost everything'. Hence, they slam the door on Pierre-Simon Laplace's 'demon', introduced in 1814. This hypothetical being has a "vast intellect", such that, with full knowledge of the state of the Universe at one time, it can completely predict the future and recall the past with no uncertainty whatsoever. Wolpert's results are particularly compelling because they are totally independent of both the details of the laws of physics and the computational characteristics of the machines.

What are the practical consequences of these findings for science? A prescription for deriving the laws of nature was proposed by Roy Frieden[4] a few years ago. He asserted that any attempt to measure a physical quantity elicits a transfer of information from the 'source' physical phenomenon to the observer. As a consequence, the observation of a physical phenomenon cannot be entirely accurate. By manipulating 'Fisher information', a measure of the quality of data, one can obtain information about the equations that govern the phenomenon in addition to the numerical value one seeks. Most fundamental equations of physics can be derived in this way (thus making physics a minor branch of statistics!). But Wolpert's work warns us that Frieden's recipe is bound to fail at least once, perhaps by producing multiple solutions in certain cases.

Another example of the relevance of Wolpert's work to science is in predicting the behaviour of chaotic systems. Through the attractor reconstruction method, in which a time series is converted into a geometrical trajectory in higher-dimensional spaces[5], one can forecast the evolution of fairly complex systems up to a specified prediction horizon. This method works quite well, and does not need the explicit knowledge of the system's governing equations (see ref. 6 for an astrophysical example). However, when one tries to infer the equations themselves, the results are often unclear or ambiguous[7], possibly as a manifestation of Wolpert's theorems.

Finally, Wolpert's findings have a bearing on the possible limitations of two theories in physics. One set of limitations concerns the standard model of particle physics, which has accumulated a long list of shortcomings[8]: these include an exceedingly high predicted cosmological constant, failure to predict the mass of the Higgs particle, and failure to account for the 'dark matter' in the Universe. The other limitation is our inability to bring quantum mechanics and gravity into a single theory, although several viable alternative theories are being studied[9]. Quantum electrodynamics, a refinement of quantum mechanics, is defined by just two parameters (the charge and mass of the electron), whereas quantum gravity would require infinitely many parameters, and hence infinite experiments to determine those parameters, making it so far a meaningless theory. It is possible, though, that these various theories, along with all that we have learned in physics and other scientific disciplines, will yet merge into the best science can do: a theory of almost everything. ■

P.-M. Binder is in the Department of Physics and Astronomy, University of Hawaii, Hilo, Hawaii 96720, USA, and at the Center for Nonlinear Dynamics, University of Texas, Austin.
e-mail: pbinder@hawaii.edu

1. Barrow, J. D. *Impossibility: The Limits of Science and the Science of Limits* (Oxford Univ. Press, 1998).
2. Wolpert, D. H. *Physica D* **237,** 1257–1281 (2008).
3. Wolfram, S. *Phys. Rev. Lett.* **54,** 735–738 (1985).
4. Frieden, B. R. *Science from Fisher Information: A Unification* (Cambridge Univ. Press, 2004).
5. Kantz, H. & Schreiber, T. *Nonlinear Time Series Analysis* (Cambridge Univ. Press, 2004).
6. Binder, P.-M., Crosson, I. J. & Cadmus, R. R. Jr *Astrophys. J.* **685,** L145–L148 (2008).
7. Crutchfield, J. P. & McNamara, B. S. *Complex Systems* **1,** 417–452 (1987).
8. Ellis, J. *Phys. World* **12** (12), 43–48 (1999).
9. Wilczek, F. *The Lightness of Being* (Basic Books, 2008).

Box 1 | Cantor, Gödel , Turing and the uncountable

In 1874, Georg Cantor published a proof of the existence of uncountable infinities. He started by labelling points in the interval [0, 1) with the countable infinite natural numbers (1, 2, 3, ...) as follows:

$a_1 = 0.d_{11}\, d_{12}\, d_{13}\, d_{14}\, d_{15}$ ... (for example, $a_1 = 0.31415926$ ... with $d_{11} = 3, d_{12}=1$, etc.)
$a_2 = 0.d_{21}\, d_{22}\, d_{23}\, d_{24}\, d_{25}$...
$a_3 = 0.d_{31}\, d_{32}\, d_{33}\, d_{34}\, d_{35}$ ...
...

All *d*s are digits between 0 and 9. There is at least one number $a_x = 0.d_{x1}d_{x2}d_{x3}$ ... in the unit interval [0, 1) such that $d_{x1} \neq d_{11}$, $d_{x2} \neq d_{22}$, and so on, guaranteeing that it differs from each number in the list by at least one digit, and hence it cannot be in the list. This proves that the number of points in the unit interval is not countable, a proof known as Cantor diagonalization.

A related result, the diagonal lemma, has played an important part in the proof of several incompleteness theorems. In 1931, Kurt Gödel proved that any mathematical system that includes enough of the theory of natural numbers contains statements that cannot be proved to be either true or false, and is thus incomplete. The general argument for the proof is based on Epimenides' liar's paradox — is 'This sentence is false' a true or a false statement? — but it replaces 'false' with 'unprovable'. The construction of a number that represents a concrete, undecidable statement requires diagonalization techniques.

In the mid-1930s, Alan Turing used similar methods to prove that no general algorithm can determine whether a given Turing machine — a computer — halts for a given input. It is thus not surprising that Wolpert's proofs also rely on diagonalization arguments. **P.-M.B.**

MOLECULAR BIOLOGY

Bound to splice

Bruce Futcher and Janet K. Leatherwood

Messenger RNAs don't usually correspond exactly to DNA — portions of the primary transcript, known as introns, are removed by splicing. A study reveals new ways in which splicing can be regulated.

Complex eukaryotes, such as animals, have extensive RNA splicing to remove sequences that don't encode proteins (introns) and to connect those that do (exons). Often, several messenger RNAs can be generated by a single gene, because different patterns of splicing place different exons into the final mRNAs. As an example, the *Dscam* gene of the fruitfly *Drosophila* contains 24 exons, and is thought to encode 38,016 different protein isoforms by alternative splicing (although we're not sure anyone has counted)[1]. In this particular case, 'docking' and 'selector' sequences within the primary transcript help regulate splicing. But in general, understanding exactly how a cell picks and chooses among the many possible combinations of splices has been a long-standing problem[2]. Reporting on page 997 of this issue, Moldón *et al.*[3] investigate how a gene's promoter region — the sequence that regulates gene expression — can regulate splicing.

As ever, one approach is to work with a simpler eukaryote. The fission yeast *Schizosaccharomyces pombe* has a sophisticated splicing apparatus, and regulates splicing, although this regulation is somewhat different from that in more complex eukaryotes. In the latter, introns are humongous, and the splicing system may therefore focus on recognizing the relatively small exons. When splicing is altered, it may fail to recognize an exon, leaving that exon out of the final product, and giving alternative splicing. In yeast, introns are tiny, and the splicing system seems to focus on recognizing the introns. Thus, when splicing is altered, the system may fail to recognize an intron, retaining the intron in the final product. Whatever the reason, the usual nature of regulated splicing in animals is alternative splicing (an altered selection of exons), whereas the usual nature of

50 & 100 YEARS AGO

50 YEARS AGO

The American Air Force lunar probe attempt on October 11 failed to achieve the full objective of a circum-lunar orbit, but demonstrated that many of the technical problems of launching and accelerating a three-stage rocket to the necessary speed have been solved. Guidance of the rocket into the correct path and precise control of the final cut-off velocity appear to have been the chief problems not yet completely mastered ... In the instrument payload of some 40 lb. were included radiation detectors, which confirmed the earlier measurements on *Explorer* IV of the intense radiation belt surrounding the Earth. Preliminary analyses of the results have shown a significant decrease in intensity beyond several Earth radii, and seem to confirm the idea that the radiation is due to trapping of cosmic particles by the Earth's magnetic field ... No doubt there are some who will not regret the failure to penetrate the mystery of the Moon's unseen face, but it must be only a matter of time until this is done ...

From *Nature* 18 October 1958

100 YEARS AGO

On October 10, in the presence of the leading aëronautical experts of France, Mr. Wilbur Wright, with M. Painlevé as a passenger, accomplished a flight of 1h. 9m. 45.6s in duration, the distance covered being estimated at nearly seventy kilometres. This successful flight is the last demanded of Mr. Wright by the French syndicate which has acquired the local rights in his aëroplane by the payment of 10,000l. at once and 10,000l. in a month's time, after three men have been trained to work the machine ... [O]n November 1 the Société navale des Chantiers de France will begin at Dunkirk the construction of fifty Wright aëroplanes, which are to be sold at the price of 1000l. each.

From *Nature* 15 October 1908

regulated splicing in yeast is intron retention[4]. Nevertheless, the mechanisms underlying these two kinds of regulated splicing may be similar.

In *S. pombe*, regulated splicing occurs in meiosis[5]. Many meiosis-specific proteins are toxic to vegetative (that is, mitotic) cells, and their expression is kept turned off by multiple mechanisms[6]. For many meiosis-specific genes, transcription is repressed in vegetative cells, and furthermore, if and when small amounts of transcript are made, their splicing is repressed; that is, introns are retained, so that no active protein is made. When the cells enter meiosis, transcription is induced, and so is splicing.

Moldón *et al.*[3] have studied the splicing of *rem1*, a meiosis-specific gene of *S. pombe*. Transcription of *rem1* is repressed in vegetative growth, and any transcript that does get made does not get spliced. On entry into meiosis, the *rem1* transcript is induced and spliced[5,7]. The major finding of Moldón *et al.*[3] is both simple and remarkable: the information that specifies meiosis-specific splicing lies entirely inside the promoter, and not in the transcribed region. For example, when the *rem1* transcript is expressed from some other promoter, it is spliced in both vegetative and meiotic cells. Conversely, when the authors used the *rem1* promoter to drive transcription of a normal vegetative gene (*cdc2*, which has four introns, and which is usually spliced in both vegetative and meiotic cells), then splicing occurred only in meiosis, and in the same temporal pattern as for the wild-type *rem1* gene.

In meiosis, the *rem1* promoter is bound by Mei4, a meiosis-specific protein belonging to the forkhead family of transcription factors. *S. pombe* has three other forkhead transcription factors, and Moldón *et al.* suggest that, in vegetative cells, the Mei4-binding sites in the *rem1* promoter are probably occupied by one of these other factors, Fkh2. When the authors deleted the *fkh2* gene from *S. pombe*, vegetative transcription of *rem1* was slightly increased, and some of this transcript was spliced. This suggests that Fkh2 represses both transcription and splicing in vegetative cells. The authors show that Mei4, which is made only in meiosis, binds to *rem1* and turns on both transcription and splicing.

Why would one forkhead transcription factor induce splicing, but not the other? On the basis of co-immunoprecipitation experiments, Moldón *et al.* found that Mei4, but not Fkh2, forms complexes with the spliceosome. The authors therefore suggest that the Mei4 transcription factor actively recruits splicing factors to the *rem1* gene and transcript, whereas Fkh2 does not. They suggest that it is this recruitment of the spliceosome by a meiosis-specific transcription factor that is responsible for meiosis-specific splicing. However, less direct explanations are also possible: transcription factors can affect the conformation of chromatin, the rate of mRNA elongation, and the 5′ capping and 3′ polyadenylation processing of transcripts, and all these are interrelated with splicing[8].

The model proposed for *rem1* is exciting, but still leaves us with a major puzzle. How can Fkh2 at a promoter inhibit splicing of *rem1* and even an unrelated gene such as *cdc2* whose transcript has good splicing signals? Maybe control of RNA processing (5′ capping and 3′ polyadenylation as well as splicing) will be a more general feature of promoters. It now seems that most of the genome is transcribed to at least some degree, so turning gene expression off completely may depend on regulating steps of RNA processing in addition to controlling efficiency of transcription.

These remarkable findings[3] leave some loose ends. First, Fkh2 regulates many vegetative transcripts, and many of these are spliced. Thus, Fkh2 is not repressing vegetative splicing at most of its targets. Second, the *rem1* promoter does not impose its usual temporal pattern of splicing when fused to the *crs1* gene[5], in contrast to the results obtained here with *cdc2* (ref. 3). Third, in the model proposed, the failure of vegetative splicing for the *rem1*-driven *cdc2* transcript seems to suggest little or no ability to splice *rem1*-driven transcripts post-transcriptionally, in contrast to *Saccharomyces cerevisiae*, in which most splicing may be post-transcriptional[9]. It should also be noted that the study of intron retention is complicated by two subtle and nasty artefacts: unlike mRNAs generated by alternative splicing, mRNAs generated by complete intron retention are perfectly co-linear with the DNA. Thus, the polymerase chain reaction following reverse transcription (RT-PCR), which is used to amplify an mRNA with retained introns, can amplify genomic DNA instead, with results that appear identical results (an artefact Moldón *et al.* addressed). Second, if a gene has an antisense transcript, this will never be spliced, but, in RT-PCR, will yield the same product as an unspliced sense transcript.

Still, these issues do not significantly undercut the core result for *rem1*: splicing depends on which transcription factor is bound to the promoter. Understanding exactly how Mei4 turns splicing on, and how Fkh2 keeps splicing turned off, will be interesting investigations for the future. In addition, it is a reminder that efforts to understand splicing by searching for signals inside the transcript have limitations, as some of the information is elsewhere. ■

Bruce Futcher and Janet K. Leatherwood are in the Department of Molecular Genetics and Microbiology, Stony Brook University, Stony Brook, New York 11794-5222, USA.
e-mail: janet.leatherwood@sunysb.edu

1. Schmucker, D. *et al. Cell* **101,** 671–684 (2000).
2. Wang, Z. & Burge, C. B. *RNA* **14,** 802–813 (2008).
3. Moldón, A. *et al. Nature* **455,** 997–1000 (2008).
4. Romfo, C. M. *et al. Mol. Cell. Biol.* **20,** 7955–7970 (2000).
5. Averbeck, N. *et al. Mol. Cell* **18,** 491–498 (2005).
6. Harigaya, Y. & Yamamoto, M. *Chromosome Res.* **15,** 523–537 (2007).
7. Malapeira, J. *et al. Mol. Cell. Biol.* **25,** 6330–6337 (2005).
8. Kornblihtt, A. R. *Curr. Opin. Cell Biol.* **17,** 262–268 (2005).
9. Tardiff, D. F. *et al. Mol. Cell* **24,** 917–929 (2006).

NANO-OPTICS

Optical antennas tuned to pitch

Lukas Novotny

Mapping out the resonant modes of optical antennas is part of a largely unexplored terrain — but not any more, as a study that applies a luminescence technique to gold antennas demonstrates.

Optical antennas are devices designed to efficiently convert optical radiation into localized energy, and vice versa. They enhance the interaction between light and matter, and hence have the potential to boost the efficiency of optoelectronic devices ranging from light-emitting diodes to solar cells. Writing in *Physical Review Letters*, Ghenuche *et al.*[1] demonstrate that the resonances of gold antennas can be mapped out spectrally and spatially with a technique known as two-photon excited luminescence (TPL) spectroscopy. The experimental mapping of optical antenna resonances by TPL not only validates theoretical models but also provides an efficient means for visualizing energetically 'hot' regions in arbitrary antenna configurations.

TPL in gold is a weak process that can be enhanced if coupled to surface plasmons[2], the size and frequency-dependent resonances of a gas of electrons. Regions of high TPL intensity are associated with strong electric fields. Therefore, a spatial map of the intensity describes an antenna's characteristic electric-field distribution. Although TPL has been used before to characterize different forms of optical antennas[3–5], Ghenuche *et al.*[1] provide the first spectrally resolved images; that is, they record spatial maps of the TPL intensity for different excitation wavelengths. The resulting data describe the spatial–spectral evolution of characteristic antenna resonances. In their experiments, the authors focused a pulsed laser beam on optical antennas fabricated by electron-beam lithography and detected the TPL emission of the antennas' response. Spatial maps of the TPL intensity were recorded pixel-by-pixel by raster scanning the antennas through the laser focus.

The optical antennas they studied are gap antennas that consist of gold nanowires separated by a narrow gap of a few nanometres. The resonance wavelength of a single gold nano-rod scales linearly with the antenna length[6]. But the resonance wavelength of a gap antenna also depends sensitively on the gap size[7]. The gap antennas of Ghenuche *et al.* exhibited a sharp resonance at an excitation wavelength, λ, of 730 nanometres, in excellent agreement with the authors' model calculations. This agreement allowed them to identify their measured antenna resonances as higher-order modes, namely as $3\lambda/2$ and $5\lambda/2$ resonances. Antennas operating at the fundamental resonance ($\lambda/2$) can be expected to be more efficient than those fabricated by Ghenuche *et al.*, but they need to be shorter, requiring more stringent fabrication tolerances[3]. Also, an optical antenna interacts with the radiation it generates and it would be interesting to measure the influence of the antenna length on the TPL emission spectrum.

The spatial resolution of spectroscopic TPL imaging is limited by diffraction to about a third of the wavelength of the illuminating laser. But although this resolution can identify the modes of an antenna, higher resolution is needed to fully resolve the details of the energy distribution along the antenna. One possible method for increasing resolution is to use a sharp tip or particle to locally scatter the TPL signal generated by an antenna[5]. The tip is raster scanned over the laser-irradiated antenna in close proximity, and the TPL intensity is measured as a function of the tip's scan coordinates. The resolution of this approach is only limited by the sharpness of the scatterer, and can be as high as 10 nanometres. However, strong interaction between the local scatterer and the antenna must be avoided, so that the antenna's behaviour is not perturbed.

The energy distribution along an optical antenna depends both on the wavelength and on the spatio-temporal properties of the exciting light. For example, the antenna's response depends on whether the excitation laser beam irradiates the entire antenna or only a fraction of it, as is the case in this new study[1]. Similarly, the antenna's response depends on the phase profile of the laser beam and on its polarization state. The use of polarization- and phase-engineered laser pulses will allow a dynamic control of the energy distribution along an optical antenna[8].

For practical applications, antennas need to be coupled to a source of radiation (transmitter) or a sink of radiation (receiver), such as an ion, molecule or quantum dot. Efficient coupling influences the overall behaviour of the antenna, so the antenna and the receiver/transmitter need to be considered as a single system. For example, a gold gap antenna coupled to an impedance-matched source at its gap becomes a half-wave antenna[6,9] — that is, an antenna that no longer feels the gap. Likewise, the presence of an antenna modifies the properties of the receiver/transmitter. For instance, for an atomic or molecular system, the antenna modifies the transition rates between the different energy states[10] and, in the case of a strong interaction, it even affects the energy-level structure. Furthermore, the highly localized fields near the antenna open up new interaction mechanisms between light and matter, such as higher-order multipole transitions or momentum-forbidden transitions. These interactions, which are inaccessible in the free space, have the potential to enrich optical spectroscopy and provide new strategies for optical sensing and detection.

The theoretical study of optical antennas is still in its infancy, and we can expect exciting results and developments in the near future. Optical antennas will boost the efficiency of light–matter interactions and enable control of photophysical processes on the nanometre scale. This has many foreseeable applications (Fig. 1), such as high-resolution microscopy and spectroscopy, photovoltaics and photocatalysis, and solid-state lighting. ■

Lukas Novotny is at the Institute of Optics, University of Rochester, Rochester, New York 14627, USA.
e-mail: novotny@optics.rochester.edu

1. Ghenuche, P. *et al. Phys. Rev. Lett.* **101,** 116805 (2008).
2. Beversluis, M. R. *et al. Phys. Rev. B* **68,** 115433 (2003).
3. Muehlschlegel, P. *et al. Science* **308,** 1607–1609 (2005).
4. Schuck, P. J. *et al. Phys. Rev. Lett.* **94,** 017402 (2005).
5. Bouhelier, A. *et al. Appl. Phys. Lett.* **83,** 5041–5043 (2003).
6. Novotny, L. *Phys. Rev. Lett.* **98,** 266802 (2007).
7. Aizpurua, J. *et al. Phys. Rev. B* **71,** 235420 (2005).
8. Aeschlimann, M. *et al. Nature* **446,** 301–304 (2007).
9. Alù, A. & Engheta, N. *Phys. Rev. Lett.* **101,** 043901 (2008).
10. Kühn, S. *et al. Phys. Rev. Lett.* **97,** 017402 (2006).

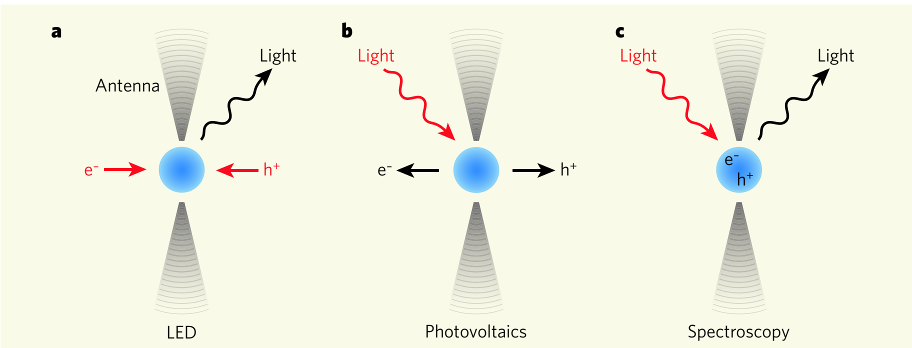


Figure 1 | Optical antennas and their applications. The gold gap antennas studied by Ghenuche *et al.*[1] can find several applications. **a,** In a light-emitting diode (LED), charge carriers — electrons (e^-) and electron holes (h^+) — are recombined in a medium (blue) to produce light. **b,** In photovoltaics, incident light causes separation of the charge carriers. **c,** In spectroscopy, incident light polarizes the medium of interest, and this polarization gives rise to outgoing radiation. In all three cases, the optical antenna (grey) enhances the efficiency of the input–output conversion process.

NATURE|Vol 455|16 October 2008

OCEANOGRAPHY

Ripples run deep

Steve Elgar

Tiny, wind-generated ripples on the sea surface can interact and produce pressure changes felt on the ocean floor. The same line of study points to a basic distinction between two types of surface wave.

The sea floor is a vibrant place — earthquakes, volcanoes, man-made explosions and ships are all causes of disturbances that can be monitored by instruments on the ocean bottom. Data from arrays of sea-floor seismometers also provide information about the structure of Earth's interior[1], but in this case in particular, intricate analysis is required to separate the desired signal from 'noise'. One source of such noise is the pressure fluctuations arising from wind-generated waves on the ocean surface. Despite the attenuation caused by the water column, the pressure signals from surface waves can reach the sea floor even in the deep ocean. A new and unexpected angle on this general phenomenon comes in a paper by Farrell and Munk just published in *Geophysical Research Letters*[2].

Even during a hurricane, it is fairly calm not far beneath the surface of the sea. The powerful swells are attenuated by the water between the surface and the bottom, with long waves decaying more slowly than those with shorter wavelengths. For example, the pressure signals from long-wavelength tides, with lengths of some 1,000 kilometres and periods of 12 hours or so, reach the floor of even the deep ocean. But the signals from ocean swells 150 metres long and with a 10-second period have lost about 97% of their energy when they arrive at 100 metres depth. And wind-generated surface ripples (10 centimetres long, with a period of less than 1 second) lose 99.99% of their energy in just a few metres of water.

Nonetheless, as Farrell and Munk[2] show, these wind-generated, thumbnail-sized ripples on the ocean surface can produce pressure signals that pass through the water column to the sea floor with little attenuation. When two such waves travel in opposite directions on the ocean surface, for instance after a sudden shift in the wind direction, they can interact nonlinearly with each other. The result can be the generation of a third wave with half the period of the wind waves, but with a much longer wavelength (Fig. 1). Given that attenuation in the water column decreases with increasing wavelength, the pressure signal from the nonlinearly generated long wave can reach the sea floor with little decay.

Farrell and Munk[2] demonstrate that nonlinear interactions between short (order 1 centimetre wavelength), small (heights of order 1 centimetre) waves travelling in opposite directions on the ocean surface produce waves with long wavelengths of some hundreds of metres. These nonlinear waves create extremely rapid pressure fluctuations, with periods of less than a tenth of a second, that are detectable on the sea floor more than 5 kilometres below the surface. These deep-sea pressure fluctuations are probably common, because winds shift from one direction to another as storms cross the ocean surface.

To reach their conclusions, Farrell and Munk combined theoretical work with analysis of observations[3] made over the course of a year with a sea-floor hydrophone at 5.5 kilometres depth near Wake Island in the North Pacific Ocean. Not only do the authors demonstrate that the pressure effects of interacting surface ripples can reach this depth, but their mathematical models also show that the observations suggest a fundamental difference between gravity and capillary waves. Gravity waves are long waves that are caused by gravitational restoring forces when the water surface is displaced, for example by wind. In contrast, capillary waves are so short that the associated curvature of the water surface is restored by surface tensile forces. The fundamental difference between them, proposed by Farrell and Munk, is that their spectral shapes — that is, the amount of wave energy associated with each wave period or wave frequency, f — decay as f^{-7} (gravity waves) and f^{-3} (capillary waves). The ocean-bottom data suggest that a transition from gravity to capillary waves occurs at periods of about a third of a second.

It is remarkable that Farrell and Munk[2] were able to estimate the separation between gravity and capillary waves from measurements made more than 5 kilometres below the surface. Although high-frequency 'noise' has been observed in ocean-bottom seismometer records for many years, an explanation was lacking until Farrell and Munk combined nonlinear wave theory with careful analysis of a nearly 20-year-old data set.

Taking their work further will require other developments, however. Studies[4–6] using extensive arrays of pressure gauges have provided measures of the periods, wavelengths and directions of the 150-metre-wavelength surface swells that cause ocean-bottom microseisms. But that has not been done for short waves. It would be difficult to deploy a sufficiently extensive array of sensors to estimate the distribution of short-wave parameters over the oceanic surface area (estimated at two to three times the water depth[2]) that contributes to the pressure fluctuations on the sea floor. The array would have to be huge in overall extent yet have sensors spaced at the centimetre scale to detect the wavelengths of interest.

Direct verification of Farrell and Munk's hypothesis will thus require innovative observational techniques, perhaps using high-resolution remote sensing of the sea surface combined with bottom-mounted pressure sensors. Meanwhile, even as they stand, their results are remarkable: they explain a complex geophysical phenomenon and will prompt many further studies. ■

Steve Elgar is at Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA.
e-mail: elgar@whoi.edu

1. Wookey, J. & Helffrich, G. *Nature* **454,** 873–876 (2008).
2. Farrell, W. E. & Munk, W. *Geophys. Res. Lett.* doi:10.1029/2008GL035008 (2008).
3. McCreery, C. S., Duennebier, F. K. & Sutton, G. H. *J. Acoust. Soc. Am.* **93,** 2639–2648 (1993).
4. Webb, S. C. & Cox, C. S. *J. Geophys. Res.* **91,** 7343–7358 (1986).
5. Herbers, T. H. C. & Guza, R. T. *J. Phys. Oceanogr.* **21,** 1740–1761 (1991).
6. Herbers, T. H. C. & Guza, R. T. *J. Geophys. Res.* **99,** 10035–10048 (1994).

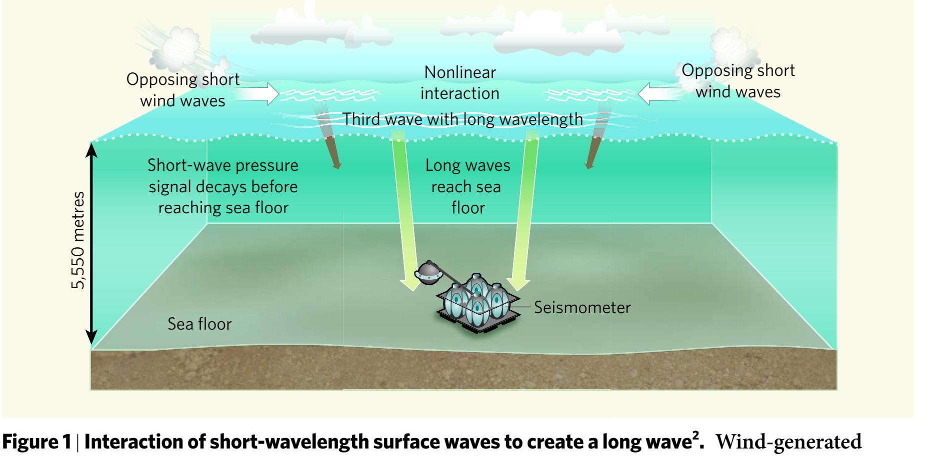
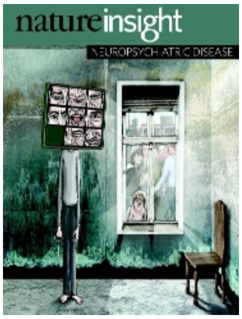


Figure 1 | Interaction of short-wavelength surface waves to create a long wave[2]. Wind-generated ripples travelling in opposite directions on the sea surface interact non-linearly with each other to generate waves with twice the period and much longer wavelengths. Unlike the pressure signal from the ripples, which decays rapidly with depth, the signal from the long-wavelength non-linear waves propagates to the sea floor with little attenuation, and is detected with ocean-bottom instruments.



Cover illustration
(Artwork by N. Spencer)

NEUROPSYCHIATRIC DISEASE

Since the time of ancient Egypt, societies have struggled to understand mental illness and to care for those affected by it. But, over the millennia, the idea that mental illness might have a biological cause arose only intermittently, and treatments ranged from the benign (exercise, humour and music) to the barbaric (exorcism, imprisonment and lobotomy). By the mid-twentieth century, however, several breakthroughs had been made. Not only did health professionals understand mental illnesses to be diseases of the brain, but a set of systematic criteria for diagnosis had been developed, together with pharmaceutical and psychological therapies that are still central to modern psychiatry.

Today, despite decades of subsequent research, the prevalence of neuropsychiatric diseases has not decreased. Our understanding of the biological mechanisms of diseases such as mood disorders, schizophrenia and autism is frustratingly limited. And, although it has long been clear that most such diseases have a strong genetic component, the identities of the genes involved have proved elusive. There is also a lack of reliable biological markers for characterizing these diseases and, perhaps unsurprisingly, treatment options are far from optimal in terms of efficacy and specificity.

There is, however, some cause for optimism. Recent advances in genomic technology and large-scale studies are helping to identify genetic variants associated with diseases. In addition, new animal models of disorders such as depression and autism are providing ways to test hypotheses about the underlying neuropathology — at the molecular, neural-circuit and behavioural levels. This Insight highlights recent successes and new ideas in this crucial area of research. The hope is that developments such as these will lead to integrative approaches for designing better therapeutic strategies.

We are pleased to acknowledge the financial support of Eli Lilly & Company in the form of an educational grant that contributed to producing this Insight. As always, *Nature* carries sole responsibility for editorial content and peer review.

I-han Chou and Tanguy Chouard, Senior Editors

INTRODUCTION

REVIEWS

nature insight

NATURE|Vol 455|16 October 2008|doi:10.1038/nature07454

A glimmer of light for neuropsychiatric disorders

Steven E. Hyman[1]

Understanding the pathogenesis of neuropsychiatric disorders is a substantial challenge for neurobiologists. It has long been hoped that identifying alleles that confer increased risk of such disorders would provide clues for neurobiological investigation. But this quest has been stymied by a lack of validated biological markers for characterizing and distinguishing the different disorders and by the genetic complexity underpinning these diseases. Now, modern genomic technologies have begun to facilitate the discovery of relevant genes.

Neuropsychiatric disorders severely compromise the well-being of those affected, with their negative effects on general health and on the ability of children to learn and of adults to work. These disorders have a relatively high prevalence[1,2], can have an early onset (for example, autism in childhood and schizophrenia in young adulthood) or a relapsing–remitting course (as in mood and anxiety disorders and obsessive–compulsive disorder), and often have disabling symptoms. For these reasons, they exert, in aggregate, a devastating impact on the families involved and on the human capital of a society. To help guide countries in assessing the investment needed in the health sector, the World Health Organization (WHO) has adapted a measure of disease burden[3], the Disability Adjusted Life Year (DALY), which is the sum of the years of life lost to premature mortality (YLL) in a population and the years lived with disability (YLD). Neuropsychiatric disorders, especially mood disorders, are the leading risk factors for suicide, but the major effect of these illnesses on disease burden is the disability they create. For the year 2000, the WHO estimated that neuropsychiatric disorders were responsible for 12% of all DALYs worldwide and 31% of YLDs[4] (Fig. 1). For North America and Europe, where infectious disease and malnutrition are far less prominent than in the developing world, neuropsychiatric disorders were responsible for a striking 43% of all YLDs[4].

Despite the disease burden attributable to neuropsychiatric disorders, and despite significant research, their mechanisms of pathogenesis and precise genetic and non-genetic risk factors have remained stubbornly out of reach. Although sustained and clever exploitation of clinical observations has produced a useful pharmacopoeia, almost all recently introduced psychotherapeutic drugs are based on reverse engineering the mechanisms of existing drugs. Arguably, no new drug targets or therapeutic mechanisms of real significance have been identified for more than four decades[5]. This parlous state of affairs is finally beginning to improve, in part through the application of new genomic technologies coupled to advances in neuroscience. The articles in this Insight explore some of these advances. In this Introduction, I discuss the obstacles that have delayed progress, look at some recent successes in neuropsychiatric genetics and touch on the promise of these findings for neurobiology.

Challenges

There are many reasons for the slow progress in understanding neuropsychiatric disorders. First, it is a daunting challenge to delineate the neurobiology of higher cognition, emotion regulation and executive function ('top-down' control of cognition and behaviour). Yet these are precisely the functions impaired in neuropsychiatric disorders. Second, it has been difficult to produce convincing and useful animal models of human cognitive, social and emotional functions and their dysfunction. Third, the living human brain is protected from the prying eyes and instruments of investigators by its bony skull, and for both practical and ethical reasons, human experimental neurobiology is mostly limited to indirect, non-invasive methods of investigation. Fourth, although neuropsychiatric disorders are highly influenced by genes, the genetic risk factors for common neuropsychiatric disorders have proved fiendishly complex[6,7]. And last, there are no well-validated objective biological markers by which to delimit precise phenotypes to use in genetics or any other type of research[8]. The lack of objective tests has meant that diagnosis has been made on the basis of phenomenological criteria (symptoms, signs and course of illness) selected by expert consensus. The current diagnoses listed in the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV)[9] are useful in a rough and ready way, and indeed necessary if clinicians and researchers are to communicate, but it would be hard to argue that they provide a mirror of nature[8].

Genetics

Despite these obstacles, neurobiology is making steady progress in discovering the molecular, cellular, synaptic, neural-systems, computational and cognitive underpinnings of brain function and its infirmities. The complexity of the task suggests, however, that progress will be painfully slow from the point of view of public health. Measures of genetic influence, such as heritability and recurrence risk ratio (Table 1), suggest that autism[10], schizophrenia[11,12] and bipolar disorder[12] are among the most 'genetic' of any common medical disorder[13]. Many other common neuropsychiatric disorders, including major depression[14], also have a significant genetic component. Therefore, it has long been hoped that alleles conferring increased disease risk could be identified. These would, in turn, provide important tools for neurobiology and would ultimately aid in developing treatments. In Alzheimer's disease, for example, the discovery of disease-causing mutations within the amyloid precursor and presenilin genes identified molecular pathways involved in pathophysiology and thus led to potential drug targets[15]. Spatial and temporal patterns of expression of risk-conferring genes can identify cells and circuits involved in illness and the timing of disease onset; antibodies raised against the protein products of implicated genes can

[1]Office of the Provost, Harvard University, Department of Neurobiology, Harvard Medical School, Massachusetts Hall, Cambridge, Massachusetts 02138, USA.

be used to probe pathogenesis; and risk-conferring genes can be used to construct mouse models of disease.

Over the past two decades, however, efforts to identify risk-conferring alleles for the common forms of neuropsychiatric disorder have largely been unrewarding. Despite the significant role for genes highlighted by aggregate measures of their influence (Table 1), the underlying genetics of common neuropsychiatric disorders has proved highly complex, as attested by unpredictable patterns of segregation in families, lack of Mendelian ratios in twin studies and serious difficulties in replicating genetic linkage studies. There is much evidence that similar neuropsychiatric symptoms can result from different combinations of genetic risk factors[7]. Conversely, the same genetic variant might be associated with multiple DSM-IV diagnoses[7,9]. For example, a balanced translocation discovered in a large Scottish family disrupts two genes on chromosome 1: disrupted in schizophrenia 1 (*DISC1*) and (*DISC2*)[16]. *DISC1* has become a well-studied positional candidate for risk of neuropsychiatric disorders, but even within the index family, the original translocation is found in individuals with different DSM-IV diagnoses: schizophrenia, schizo-affective disorder and recurrent major depression[16]. Other forms of sequence variation within *DISC1* have subsequently been associated with schizophrenia, schizo-affective disorder, bipolar disorder, major depression and both broad and narrow diagnoses of autism[17].

At one extreme, heterogeneous genetic pathways to shared phenotypes have been shown to result from interactions among different common genetic variants at multiple genomic loci, often described as the 'common disease–common variant hypothesis'. At the other extreme, they can result from diverse rare, highly penetrant mutations, with each individual mutation causing disease in a small number of families (as is the case for retinitis pigmentosa[18]). The variants that contribute to common diseases are typically single-nucleotide polymorphisms (SNPs) that are not individually deleterious but produce risk through interaction with variants at other loci and with non-genetic factors[19].

Rare, highly penetrant disease-causing mutations are by definition deleterious, causing abnormalities by loss or gain of RNA or protein functions. Not all such mutations are fully penetrant, depending on their precise biology and, potentially, on other genetic and environmental factors (such as the avoidance of phenylalanine in phenylketonuria). The identification of rare, highly penetrant disease-causing alleles is relatively straightforward with modern technologies. When common variants produce large effects on risk, they are also fairly easily identified, as in the case of the effect of the apolipoprotein E variant allele *APOE* ε4 in late-onset Alzheimer's disease[20] and a variant gene encoding complement factor H in age-related macular degeneration[21]. The identification of common risk-conferring variants has, however, proved extremely challenging in most cases, because of their relatively small contribution to the disease phenotype (typically a 1.1–1.3-fold increase in risk)[19]. Thus, the search for common variants that contribute to the risk of common disorders, such as type 2 diabetes[22], inflammatory bowel disease and breast cancer[23], has only begun to be successful with the advent of tools such as the human haplotype map of SNPs and genome-wide association studies using very large populations or meta-analyses[19,22].

In some cases, individual SNPs can produce deleterious mutations, especially when they influence protein translation or messenger RNA splicing. Copy-number variation (CNV), which involves larger structural alterations in chromosomes, including deletions, insertions, duplications and more complex genomic rearrangements, is more likely to produce biological effects by disrupting one or more genes or by quantitatively altering gene expression. In addition to harbouring millions of SNPs, the human genome harbours substantial CNV, some of it occurring *de novo* rather than being inherited. Even when risk-conferring variants are found and validated, the task of relating them to pathophysiology or to predictive testing has only just begun. Large-scale epidemiological studies are needed to better define the relationship of gene variants to particular phenotypes. Genetic association studies identify risk-conferring alleles independently of knowledge of their biological roles, so there may be little information to guide subsequent laboratory investigations. Moreover, because of the diverse genetic, environmental and random factors that lead to these common diseases, what we think of as a single 'disease' is not in a strict sense a homogeneous entity for which there is a 'Platonic' ideal phenotype. Common diseases are more likely to represent families of diseases that share major pathophysiological and symptomatic features but can differ in important characteristics such as age of onset, severity of symptoms, rate of progression, and response to treatment[24].

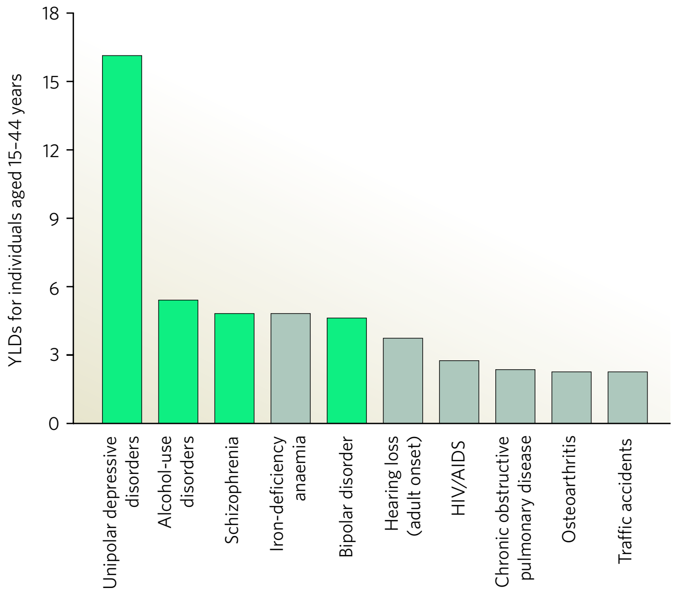


Figure 1 | Top ten causes of disability worldwide for individuals aged 15–44, as estimated for the year 2000. The contribution of each condition is quantified as years lived with disability (YLDs). Neuropsychiatric conditions are shown in green. Data are taken from ref. 4.

Phenotyping and clinical diagnosis

Untangling the genetic causes of neuropsychiatric conditions has proved difficult, not only because of intrinsic genetic complexity but also because of the lack of objective diagnostic tests. Despite a decades-long search for objective markers or for intermediary neurobiological phenotypes, such as patterns of cortical thinning in schizophrenia or stress-hormone secretion in depression, validated and widely accepted markers still remain more of a goal than a reality. By contrast, even though common conditions such as type 2 diabetes and hypertension are also heterogeneous and have had shifting diagnostic criteria over time[25], their diagnoses are based on objective tests that limit the heterogeneity of population samples used in research.

The DSM-IV diagnoses of conditions such as autism, schizophrenia, bipolar disorder, major depression and obsessive–compulsive disorder are not entirely without external validation: the high heritabilities of many of these disorders, and the remarkable cross-cultural similarity of symptoms, argues that the DSM-IV diagnosis is picking up a real signal. Of course, the editors of the DSM-IV did not ignore possible objective markers but had to develop diagnostic criteria in their absence. The reliance of the DSM-IV on clinical phenomenology means, however, that many patients do not fit any single diagnosis and that there are many anomalous observations, such as symptoms of schizophrenia and diverse mood disorders segregating within the same families[7,26].

A complicating factor for phenotyping in genetic and neurobiological studies is that the DSM-IV classification rests on an arbitrary decision to treat all disorders as 'categorical' rather than considering some to represent quantitative deviations from normality. The DSM-IV thus posits clear boundaries between 'ill' and 'well' and between different disorders. There is, however, evidence that many neuropsychiatric disorders, such as autism, schizophrenia[11] and major depression[27], are better modelled quantitatively on multiple symptom dimensions, by analogy with hypertension. A consequence of the sharp boundaries drawn between 'ill' and 'well' is seen in family studies of schizophrenia: relatives of individuals

Table 1 | Relative risk of selected disorders

Disorder	Prevalence in the population (%)	Relative risk, λ_1*	Heritability†
Autism (narrow)	0.2	25	0.9
Autism (broad)	0.7	25	0.9
Schizophrenia	1.0	9	0.8–0.9
Bipolar disorder	1.0	8	0.7–0.8
Major depression	17	2–5	0.35

*λ_1 is a recurrence risk ratio, in this case the fold increase in risk of developing the disorder over population base rates for a first-degree relative of an affected individual. †Heritability is the proportion of phenotypic variation attributable to genetic variation.

with schizophrenia, who are considered well by DSM-IV criteria, may show cortical thinning in structural neuroimaging studies and may suffer from cognitive abnormalities[7,11]. A consequence of the sharp boundaries between disease categories is that many patients receive multiple DSM-IV diagnoses (described as co-morbidity), such as generalized anxiety disorder and major depression, even when the symptoms may be aetiologically related and may respond to the same treatment.

As work begins on DSM-V, the lack of objective markers remains a major problem. That said, there are proposals that might help the DSM-V fit better with scientific data. Such proposals include the addition of quantitative dimensions to the manual for some disorders, based on data derived from sources that include epidemiology, family studies and genetics. Other proposals include the possibility of updating diagnostic criteria for specific disorders without having to wait for an entirely new edition. It is probable that improvements in diagnosis will involve iterative processes in which, for example, genetic and neurobiological discoveries might suggest better disease groupings, which, in turn, might lead to additional genetic and neurobiological advances.

Early progress and future directions

Despite the challenges posed by the DSM-IV classification, as well as by a lack of objective markers for phenotyping and by genetic complexity, this year has seen significant progress in finding common variants that might contribute to bipolar disorder and schizophrenia and in finding rare, highly penetrant mutations that might give rise to schizophrenia and autism. The studies that searched for common variants contributing to bipolar disorder[28] and schizophrenia[29] assembled large numbers of subjects through the work of international consortia. Large sample sizes are needed because of the heterogeneity of the patients and because of the modest relative risk typically contributed by common variants. The study of bipolar disorder[28] achieved significant associations with the genes *ANK3* and *CACNA1C* only after combining two previously published studies and adding a new one. The ANK3 protein has a role in the assembly of voltage-gated Na^+ channels; *CACNA1C* encodes the α1c subunit of the L-type Ca^{2+} channel. Interestingly, Na^+ channels are targets of the anticonvulsant drugs used to treat bipolar disorder. In the study of schizophrenia[29], the clearest evidence for association with a gene known as *ZNF804A* required a meta-analysis. Strikingly, the strength of an association with *ZNF804A* increased when the affected schizophrenia phenotype was extended to include bipolar disorder. If this finding holds up, it bears witness either to the pleiotropic effects of genes in neuropsychiatric disorders or to a pressing need to revise the DSM-IV with respect to its categorical split between schizophrenia and bipolar disorder, a separation that dates back to the psychiatrist Emil Kraepelin (1856–1926).

The strategy of searching for highly penetrant disease-causing mutations has received much attention in neurology but perhaps too little in psychiatry, despite the fact that single-gene disorders with autistic or psychotic symptoms have long been known. Rett syndrome[30], fragile X syndrome[31] and tuberous sclerosis[32] include symptoms of autism. Chromosome 22q11.2 deletions associated with DiGeorge syndrome and velocardiofacial syndrome produce psychosis in 30% of patients[33,34]. Historically, there had been a concern that study of these rare disorders would not shed light on the common neuropsychiatric disorders with which they shared similar symptoms. As advances in genetics have demonstrated the aetiological heterogeneity of diseases such as autism[35] and schizophrenia[7,11], it has become clear that useful clues should be seized wherever they arise. Just as several neurodegenerative diseases, including Alzheimer's disease[15,36], Parkinson's disease[37] and amyotrophic lateral sclerosis[38], can result either from multiple common variants or from rare, highly penetrant mutations, it now seems that both schizophrenia[29,39–41] and autism[35,42,43] show both types of genetic risk as well. Indeed a spate of recent papers has correlated excess CNV with risk of schizophrenia[39–41]; others have closed in on mutations that might cause autism by exploiting methods to identify CNV[42,43]. Undoubtedly, there is also a wide variety of intermediate situations in which relatively penetrant mutations act against a background of multiple disease-modifier alleles to influence age of onset or to produce milder or more severe disease.

In this Insight, Edwin Cook and Stephen Scherer describe how the widespread CNV in the human genome[44] might contribute to neuropsychiatric disorders, while reminding us of the difficulty of proving that a particular mutation has a role in the disease phenotype of interest (see page 919). This point deserves to be emphasized. Biologically interesting candidate genes, in the absence of statistical demonstration of a causal role, run the risk of creating a diversion of effort with which this field is only too well acquainted. That said, appropriately validated, highly penetrant mutations provide significant opportunities for neurobiological investigation. Because common disease-related variants contribute only small increments of risk, they are likely to be highly sensitive to genetic background and environmental factors, whether their effects are being studied in people with the disease or in transgenic mouse models. Rare disease-causing mutations are more likely to produce robust phenotypes in mouse models, as they have in models of Alzheimer's disease[45] and Rett syndrome[46], even if they cannot be said to perfectly mimic the human disease. Understanding the biology of neuropsychiatric disorders will demand the assembly of information from multiple sources and levels of analysis.

Also in this Insight, Thomas Südhof describes the kind of compelling neurobiological analysis that is possible when penetrant mutations are identified in genes known to have significant functions in the nervous system (see page 903). The fascinating story of how mutations in the neuroligin and neurexin genes might alter the balance between excitatory and inhibitory neurotransmission is a major step forward. It is also a reminder of how far neurobiology must progress before we can understand how such mutations can produce specific deficits in social cognition. Melissa Ramocki and Huda Zoghbi focus on mutations that illuminate the importance of the regulation of gene and protein expression in the pathogenesis of neuropsychiatric disorders that begin during development (see page 912). They adduce several examples, including Rett syndrome (for which Zoghbi and colleagues were the first to identify a mutation in the *MECP2* gene[47]) and fragile X syndrome, and put forward the hypothesis that loss or gain of protein and RNA functions cause remarkably overlapping cognitive and neuropsychiatric effects because of an effect on neuronal network homeostasis. In addition, Vaishnav Krishnan and Eric Nestler illustrate the sophistication required in the neurobiological investigation of what may be the most common serious neuropsychiatric disorder — depression (see page 894). They describe the utility of both genetic animal models and environmentally produced animal models.

What is the significance for neurobiology of these recent advances in neuropsychiatric genetics? Even variants that contribute a small increment of risk, as is likely to be the case for *ANK3* and *CACNA1C*[28], can point to pathways that might be involved in pathophysiology and can thus suggest biological hypotheses and possible drug targets. Where highly penetrant mutations can be identified, they can provide particularly powerful tools for neurobiology, as illustrated by the neurexin and neuroligin story or the investigations into fragile X syndrome and Rett syndrome that have produced animal models of treatment. There is, of course, still a long distance to travel in the identification of risk-conferring alleles, of which there will be many, and in the replication of association studies and demonstration of both statistical and biological significance. There is much challenging neurobiology

to come, but it is biology that will ultimately have enormous significance for human health. Neuropsychiatric genetics has had many well-publicized false starts[48]. Perhaps not all the findings discussed here will hold up, but I feel real optimism brought about by the intelligent application of modern genomic technologies. Maybe the current glimmer of light will lead to a new dawn for the diagnosis and treatment of neuropsychiatric disorders. ■

1. Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R. & Walters, E. E. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch. Gen. Psychiatry* **62,** 617–627 (2005).
2. Kessler, R. C. *et al*. Prevalence and treatment of mental disorder, 1990 to 2003. *N. Engl. J. Med.* **352,** 2515–2513 (2005).
3. Murray, C. J. & Lopez, A. D. *The Global Burden of Disease* (Harvard Univ. Press, 1996).
4. World Health Organization. *World Health Report 2001* (WHO, 2001).
5. Nestler, E. J., Hyman, S. E. & Malenka, R. J. *Molecular Neuropharmacology: Foundation for Clinical Neuroscience* 2nd edn (McGraw-Hill, 2008).
6. Kendler, K. S., Davis, C. G. & Kessler, R. C. The familial aggregation of common psychiatric and substance use disorders in the National Comorbidity Survey: a family history study. *Br. J. Psychiatry* **170,** 541–548 (1997).
7. Craddock, N., O'Donovan, M. C. & Owen, M. J. The genetics of schizophrenia and bipolar disorder: dissecting psychosis. *J. Med. Genet.* **42,** 193–204 (2005).
8. Hyman, S. E. Can neuroscience be integrated into the DSM-V? *Nature Rev. Neurosci.* **8,** 725–732 (2007).
9. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* 4th edn, text revision (American Psychiatric Association, 2000).
10. Freitag, C. M. The genetics of autistic disorders and its clinical relevance: a review of the literature. *Mol. Psychiatry* **12,** 2–22 (2007).
11. Sullivan, P. F., Kendler, K. S. & Neale, M. C. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch. Gen. Psychiatry* **60,** 1187–1192 (2003).
12. Merikangas, K. R. & Risch, N. Will the genomics revolution revolutionize psychiatry? *Am. J. Psychiatry* **160,** 625–635 (2003).
13. Merikangas, K. R. & Risch, N. Genomic priorities and public health. *Science* **302,** 599–601 (2003).
14. Kendler, K. S., Gatz, M., Gardner, C. O. & Pederson, N. L. A Swedish national twin study of lifetime major depression. *Am. J. Psychiatry* **163,** 109–114, (2006).
15. Hardy, J. A hundred years of Alzheimer's disease research. *Neuron* **52,** 3–13 (2006).
16. St. Clair, D. *et al.* Association within a family of a balanced autosomal translocation with major mental illness. *Lancet* **336,** 13–16 (1990).
17. Hennah, W. *et al. DISC1* association, heterogeneity and interplay in schizophrenia and bipolar disorder. *Mol. Psychiatry* advance online publication, doi:10.1038/mp.2008.22 (4 March 2008).
18. Pacione, L. R., Szego, M. J., Ikeda, S., Nishina, P. M. & McInnes, R. R. Progress toward understanding the genetic and biochemical mechanisms of inherited photoreceptor degenerations. *Annu. Rev. Neurosci.* **26,** 657–700 (2003).
19. Hunter, D. J., Altshuler, D. & Rader, D. J. From Darwin's finches to canaries in the coal mine — mining the genome for new biology. *N. Engl. J. Med.* **358,** 2760–2763 (2008).
20. Strittmatter, W. J. *et al.* Apoliprotein E: high avidity binding to β-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc. Natl Acad. Sci. USA* **90,** 1977–1981 (1993).
21. Edwards, A. O. *et al.* Complement factor H variant increases the risk of age-related macular degeneration. *Science* **308,** 419–421 (2005).
22. Zeggini, E. *et al.* Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nature Genet.* **40,** 638–645 (2008).
23. Easton, D. F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447,** 1087–1093 (2007).
24. Ciardiello, F. & Tortora, G. EGFR antagonists in cancer treatment. *N. Engl. J. Med.* **358,** 1160–1174 (2008).
25. Chobanian, A. V. *et al.* National High Blood Pressure Education Program Coordinating Committee. Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Hypertension* **42,** 1206–1252 (2003).
26. Millar, J. K. *et al.* Disruption of two novel genes by a translocation co-segregating with schizophrenia. *Hum. Mol. Gen.* **9,** 1415–1423 (2000).
27. Kendler, K. S. & Gardner, C. O. Jr. Boundaries of major depression: an evaluation of DSM-IV criteria. *Am. J. Psychiatry* **155,** 172–177 (1998).
28. Ferreira, M. A. R. *et al.* Collaborative genome-wide association analysis supports a role for *ANK3* and *CACNA1C* in bipolar disorder. *Nature Genet.* **40,** 1056–1058 (2008).
29. O'Donovan, M. C. *et al.* Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nature Genet.* **40,** 1053–1055 (2008).
30. Chahrour, M. & Zoghbi, H. Y. The story of Rett syndrome: from clinic to neurobiology. *Neuron* **56,** 422–437 (2007).
31. Reddy, K. S. Cytogenetic abnormalities and fragile-X syndrome in autism spectrum disorders. *BMC Med. Genet.* **6,** 3–19 (2005).
32. Lewis, J., Thomas, H., Murphy, K. & Sampson, J. Genotype and psychological phenotype in tuberous sclerosis. *J. Med. Genet.* **41,** 203–207 (2004).
33. Williams, N. M., O'Donovan, M. C. & Owen, M. J. Chromosome 22 deletion syndrome and schizophrenia. *Int. Rev. Neurobiol.* **73,** 1–27 (2006).
34. Kobrynski, L. J. & Sullivan, K. E. Velocardiofacial syndrome, DiGeorge syndrome: the chromosome 22q11.2 deletion syndromes. *Lancet* **370,** 1443–1452 (2007).
35. Veenstra-Vanderweele, J., Christian, S. L. & Cook, E. H. Jr. Autism as a paradigmatic complex genetic disorder. *Annu. Rev. Genomics Hum. Genet.* **5,** 379–405 (2004).
36. Waring, S. C. & Rosenberg, R. N. Genome-wide association studies in Alzheimer disease. *Arch. Neurol.* **65,** 329–334 (2008).
37. Farrer, M. J. Genetics of Parkinson disease: paradigm shifts and future prospects. *Nature Rev. Genet.* **7,** 306–318 (2006).
38. Valdmanis, P. N. & Rouleau, G. A. Genetics of familial amyotrophic lateral sclerosis. *Neurology* **70,** 144–152 (2008).
39. Walsh, T. *et al.* Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320,** 539–543 (2008).
40. Xu, B. *et al.* Strong association of *de novo* copy number mutations with sporadic schizophrenia. *Nature Genet.* **40,** 880–885 (2008).
41. The International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455,** 237–241 (2008).
42. Weiss, L. A. *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* **358,** 667–675 (2008).
43. Morrow, E. M. *et al.* Identifying autism loci and genes by tracing recent shared ancestry. *Science* **321,** 218–223 (2008).
44. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444,** 444–454 (2006).
45. Gotz, J. & Ittner, L. M. Animal models of Alzheimer's disease and frontotemproal dementia. *Nature Rev. Neurosci.* **9,** 532–544 (2008).
46. Watase, K. & Zoghbi, H. Y. Modelling brain diseases in mice: the challenges of design and analysis. *Nature Rev. Genet.* **4,** 296–307 (2003).
47. Amir, R. E. *et al.* Rett syndrome is caused by mutations in X-linked *MECP2*, encoding methyl-CpG binding protein 2. *Nature Genet.* **23,** 185–188 (1999).
48. Egeland, J. A. *et al.* Bipolar affective disorders linked to DNA markers on chromosome 11. *Nature* **325,** 783–787 (1987).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Correspondence should be addressed to the author (seh@harvard.edu).

NATURE|Vol 455|16 October 2008|doi:10.1038/nature07455

The molecular neurobiology of depression

Vaishnav Krishnan[1,2] & Eric J. Nestler[1,2,3]

Unravelling the pathophysiology of depression is a unique challenge. Not only are depressive syndromes heterogeneous and their aetiologies diverse, but symptoms such as guilt and suicidality are impossible to reproduce in animal models. Nevertheless, other symptoms have been accurately modelled, and these, together with clinical data, are providing insight into the neurobiology of depression. Recent studies combining behavioural, molecular and electrophysiological techniques reveal that certain aspects of depression result from maladaptive stress-induced neuroplastic changes in specific neural circuits. They also show that understanding the mechanisms of resilience to stress offers a crucial new dimension for the development of fundamentally novel antidepressant treatments.

About one in six individuals in the United States will succumb to clinical depression during their lifetime[1]. Core symptoms include depressed mood, anhedonia (reduced ability to experience pleasure from natural rewards), irritability, difficulties in concentrating, and abnormalities in appetite and sleep ('neurovegetative symptoms')[2]. In addition to mortality associated with suicide, depressed patients are more likely to develop coronary artery disease and type 2 diabetes[3]. Depression also complicates the prognosis of a host of other chronic medical conditions[4,5]. The chronic, festering nature of depression contributes substantially to the global burden of disease and disability.

Despite the prevalence of depression and its considerable impact, knowledge about its pathophysiology is rudimentary compared with knowledge of other common chronic and potentially fatal multifactorial conditions, such as type 2 diabetes (Table 1). There are several explanations for this discrepancy. First and foremost, observing pathological changes within the brain remains markedly more difficult than for all other organs. Available techniques to document the aberrant function of brain circuits depend on either post-mortem studies, which have numerous limitations, or neuroimaging techniques, which rely on detecting changes in neuronal activity by using indirect markers of activation[6]. Although these approaches have provided important insights into candidate brain regions, simple increases or decreases in regional brain activity are probably insufficient to explain the complex array of symptoms caused by depression. Several animal models have also informed knowledge of the neural circuitry of depression, but there are important challenges to how information gained from these models should be interpreted (Box 1).

Second, most depression occurs idiopathically, and the limited understanding of its aetiology is reflected as a list of risk factors, such as stressful life events, endocrine abnormalities (hypothyroidism and hypercortisolism), cancers (such as pancreatic adenocarcinoma and breast tumours) and side effects of drugs (for example, isotretinoin for acne, and interferon-α for hepatitis C), among many others[2,4,7]. Genetic association studies have not uncovered strong and consistent genetic risk modifiers[8], perhaps because of the sheer heterogeneity of depressive syndromes[2,9]. Thus, genuine 'depression genes', which can be used to generate disease models in mice (for example, those for Rett syndrome or familial Alzheimer's disease), have not yet been identified. Genetic predispositions are thought to interact with environmental risk factors, such as stressful life events, which can initiate depressive episodes in some patients[10]. Still, the tendency to live in high-stress environments might also be partly heritable (as is the case for 'risk or sensation seekers')[11], emphasizing the strong genetic contribution to even 'environmentally precipitated' depressive episodes.

The official diagnosis of depression is subjective and rests on the documentation of a certain number of symptoms that significantly impair functioning for a certain duration[2]. These diagnostic criteria overlap with other conditions such as anxiety disorders, which have substantial co-morbidity with depression[12,13]. Therefore, two 'depressed' patients might have only one symptom in common[7], and a manic episode in one patient — even later in life — switches the diagnosis to bipolar disorder, which is presumably a distinct pathophysiological entity. This symptom-based diagnostic approach poses obvious obstacles to the interpretation of genome-wide association studies, as well as neuroimaging and post-mortem investigations.

In this Review, we summarize the current state of knowledge of the neural and molecular mechanisms of depression. We focus on key leading hypotheses in the field, and examine their strengths and weaknesses critically in the light of recent preclinical and translational studies. We also highlight new insights that promise to extend the understanding of depression and improve its treatment.

Neural circuitry of depression

Several brain regions and circuits regulate emotion, reward and executive function, and dysfunctional changes within these highly interconnected 'limbic' regions have been implicated in depression and antidepressant action[14] (Fig. 1). A large body of post-mortem[7,15] and neuroimaging[7,16] studies of depressed patients have reported reductions in grey-matter volume and glial density in the prefrontal cortex and the hippocampus, regions thought to mediate the cognitive aspects of depression, such as feelings of worthlessness and guilt. However, the published findings are not consistent and are often complicated by co-morbid diagnoses and medication history, and there has been limited success in demonstrating any clear cause–effect relationships of these pathological changes.

In contrast to structural studies, experiments assessing brain function, such as functional magnetic resonance imaging (fMRI) or positron-emission tomography (PET), show that activity within the amygdala and subgenual cingulate cortex (Cg25, a subregion of prefrontal cortex) is strongly correlated with dysphoric emotions: indices of neuronal activity

[1]Department of Psychiatry and [2]Department of Neuroscience, The University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. [3]Fishberg Department of Neuroscience, Mount Sinai School of Medicine, New York, New York 10029, USA.

Table 1 | A systematic comparison of major depression and type 2 diabetes

Criterion	Major depressive disorder	Type 2 diabetes
Lifetime risk	1 in 6	1 in 3
Diagnosis and monitoring	Subjective–qualitative: patients must show a depressed mood or anhedonia, as well as assorted other symptoms, for at least 2 weeks, and these symptoms must disrupt normal social and occupational functioning Patients monitored through standardized questionnaires	Objective–quantitative: diagnosis requires demonstration of an increased amount of serum glucose with classical signs (polyuria, polydipsia, obesity) or abnormal glucose tolerance (reflecting insulin resistance) Significant increases in HbA1C, a glycosylated haemoglobin, demonstrate long-standing poor glycaemic control
Aetiology and risk factors	Stressful life events (such as loss of loved ones or financial or professional crises) Genetic risk (heritability ≈ 40%) Disease genes unknown; can be idiopathic, a side effect of a drug (such as interferon-α or isotretinoin) or secondary to systemic illness (such as Cushing's syndrome or stroke, among many others)	Lifestyle factors (sedentary lifestyle, high-fat diet) Genetic risk (heritability ≈ 35%) Established disease genes (such as *PPARG*, *TCF7L2* or *KCNJ11*); can be iatrogenic (such as treatment with glucocorticoids or phenytoin)
Treatments	Monoamine reuptake inhibitors (such as tricyclic drugs, SSRIs, NRIs or SNRIs) Monoamine oxidase inhibitors (such as tranylcypromine) 'Atypical' agents (such as bupropion or mirtazapine) Electroconvulsive seizures Psychotherapy Deep brain stimulation Exercise promotes recovery	Insulin Sulphonylureas (such as tolbutamide) Meglitinides (such as repaglinide) PPAR-γ agonists (such as rosiglitazone) Biguanides (such as metformin) Glucosidase inhibitors (such as miglitol) Incretin (GLP1) mimetics (such as exenatide) Lifestyle changes (such as weight loss or exercise)
Pathogenesis	Abnormal activity of the HPA axis (hypercortisolism or hypocortisolism)? Alterations in neurotrophic signalling? Abnormal hippocampal neurogenesis? Deficits in brain reward processing? Abnormal cognitive styles (negative thinking)?	Obesity, sedentariness and genetic predisposition promote peripheral insulin resistance leading to pancreatic β-cell hyperplasia β-Cell dysfunction and failure ensues, leading to impaired glucose tolerance End-organ damage (nephropathy, neuropathy and angiopathy) occurs secondarily to hyperglycaemia, excessive protein glycation and aberrant intracellular signalling

GLP1, glucagon-like peptide 1; HbA1C, haemoglobin A1 C; HPA, hypothalamic–pituitary–adrenal axis; *KCNJ11*, potassium inwardly rectifying channel J11 gene; NRIs, selective noradrenaline reuptake inhibitors; *PPARG*, peroxisome-proliferator-activated receptor-γ gene; SNRI, serotonin–noradrenaline reuptake inhibitor; *TCF7L2*, transcription factor 7 like 2 gene.

within these regions are increased by transient sadness in healthy volunteers and are chronically increased in depressed individuals, reverting to normal levels with successful treatment[7,13]. Inspired by these findings, it was shown that deep brain stimulation applied to the white matter tracts surrounding Cg25 produced a sustained remission of depressive symptoms in a small cohort of treatment-resistant patients (patients who failed to respond to several standard treatments)[17]. Deep brain stimulation, achieved through the stereotactic surgical placement of stimulating electrodes, has also provided an acute ameliorative effect on clinical ratings when applied to the nucleus accumbens (NAc)[18], a striatal subregion that is important for reward and for hedonic deficits in depression[19].

These forebrain networks are significantly modulated by monoamine projections from midbrain and brainstem nuclei (dopamine from the ventral tegmental area (VTA), serotonin from the dorsal raphe located in the periaqueductal grey area, and noradrenaline from the locus coeruleus). In addition to controlling alertness and awareness, these neurotransmitters modulate the salience of emotional stimuli. More recent studies have investigated the role of specific hypothalamic nuclei in mediating the neurovegetative signs of depression. However, we add a note of caution: although depressive symptoms are probably mediated by dysfunction in a diffuse series of neural networks, the field has often used a simplistic 'localization of function' approach to examine limbic substrates (for example, amygdala ≈ 'fear and anxiety', NAc ≈ 'reward'). Such artificial distinctions are of limited heuristic value and reflect limitations in the ability of current technologies to understand systems-level dysfunction.

The role of monoamines

The 'monoamine hypothesis' of depression, which posits that depression is caused by decreased monoamine function in the brain, originated from early clinical observations[14,20]. Two structurally unrelated compounds developed for non-psychiatric conditions, namely iproniazid and imipramine, had potent antidepressant effects in humans and were later shown to enhance central serotonin or noradrenaline transmission. Reserpine, an old antihypertensive agent that depletes monoamine stores, produced depressive symptoms in a subset of patients. Today's antidepressant agents offer a better therapeutic index and lower rates of side effects for most patients, but they are still designed to increase monoamine transmission acutely[14], either by inhibiting neuronal reuptake (for example, selective serotonin reuptake inhibitors (SSRIs) such as fluoxetine) or by inhibiting degradation (for example, monoamine oxidase inhibitors such as tranylcypromine). Although these monoamine-based agents are potent antidepressants[21], and alterations in central monoamine function might contribute marginally to genetic vulnerability[8,22], the cause of depression is far from being a simple deficiency of central monoamines. Monoamine oxidase inhibitors and SSRIs produce immediate increases in monoamine transmission, whereas their mood-enhancing properties require weeks of treatment. Conversely, experimental depletion of monoamines can produce a mild reduction in mood in unmedicated depressed patients, but such manipulations do not alter mood in healthy controls[23]. Moreover, studies with rodent stress models have shown that enhancements in dopamine and noradrenaline transmission can have maladaptive roles in stress-related disorders by strengthening memories of aversive life events[24,25].

It is now thought that acute increases in the amount of synaptic monoamines induced by antidepressants produce secondary neuroplastic changes that are on a longer timescale and involve transcriptional and translational changes that mediate molecular and cellular plasticity[2,20]. As one example, the serotonin 5-HT_{1B} receptor interacts with a calcium-binding protein named p11, which was upregulated in cerebral cortex on chronic treatment with SSRIs and was also found to be downregulated in post-mortem cingulate cortex samples from depressed individuals[26]. The brain-specific transgenic overexpression of p11 produced an antidepressant phenotype, implicating this SSRI-mediated upregulation of p11 as an important mechanism downstream of serotonin receptor activation. Chronically administered antidepressants have also been shown to

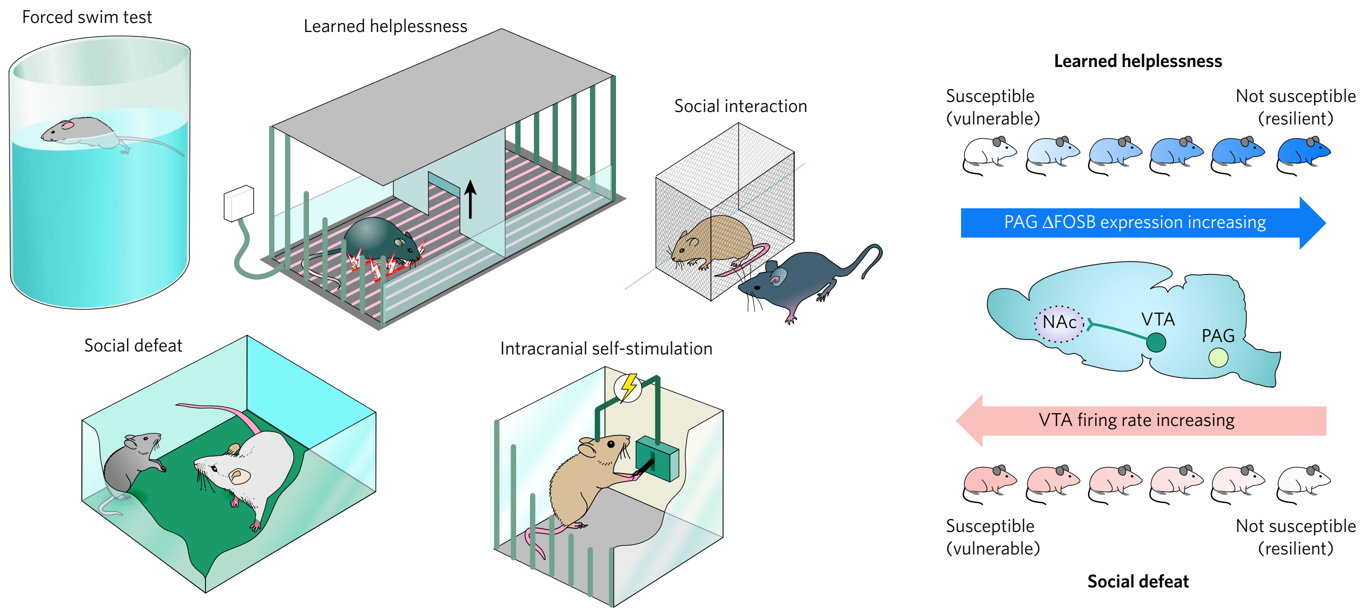
upregulate the transcription factor CREB (cyclic-AMP-response-element-binding protein), which is downstream of several serotonin and other stimulatory G-protein-coupled receptors, in the hippocampus; this effect has been validated in human post-mortem tissue and directly linked to antidepressant-like responses in animal models[2,20]. By contrast, stress activation of CREB in NAc triggers depression-like responses, which underscores crucial region-specific actions of neurotransmitters and their downstream effectors that have not been incorporated into simplistic deficiency models[19].

Monoamine-based antidepressants remain the first line of therapy for depression, but their long therapeutic delays and low (about 30%) remission rates[21] have encouraged the search for more effective agents[14,27]. The serotonin receptors involved in the action of SSRIs remain unknown, although selective agonists of the serotonin 5-HT_4 receptor produce rapid antidepressant effects in rodents (three to four days)[28]. Experiments on mice deficient in P-glycoprotein, a molecule in the blood–brain barrier that transports numerous drugs back into the bloodstream, have shown that several antidepressant agents, including the SSRI citalopram, are substrates for P-glycoprotein. Human polymorphisms in the gene encoding P-glycoprotein significantly alter antidepressant efficacy in depressed individuals[29], suggesting the value of such a pharmacogenetic approach when selecting antidepressant agents[30].

Neurotrophins and neurogenesis

Volumetric decreases observed in the hippocampus and other forebrain regions in subsets of depressed patients have supported a popular hypothesis for depression involving decrements in neurotrophic factors — neurodevelopmentally expressed growth factors that also regulate plasticity within adult brain[31,32]. These studies have focused

Box 1 | Preclinical models



Animal models of depression are evaluated for their aetiological validity: to be valid, depression-like behaviours need to be caused by the same aetiologies that trigger human depression. This is a challenging requirement, given the absence of definitive aetiologies for human depression. Current models gauge an animal's 'depression-related' responses to acute or chronic inescapable stress. These include the forced-swim test[35,53,73], which quantifies immobility in a water bath (proposed to be analogous to 'behavioural despair' observed in depression; see box figure, left). Other assays include measuring social interaction (decreases in which may model social withdrawal of depression-related conditions)[25,39,79,95], the learned-helplessness test (which measures the development of passive responses to inescapable foot shock)[84], and intracranial self-stimulation, an operant measure of the effort that an animal expends to stimulate brain reward circuits electrically[14,23]. Several of these show pharmacological validity — that is, they are sensitive to acutely administered known antidepressant compounds[14,99] — which permits the rapid screening of potentially new therapeutic agents. However, because more than half of all depressed individuals do not respond fully to available antidepressants[21], the requirement for pharmacological validity is a circular argument that deserves reconsideration. Models that use an acute stress (for example, forced swimming) are better thought of as 'tests' of coping behaviour, and are limited in their ability to recapitulate a long-lived multidimensional syndrome such as depression. Efforts to create the latter are almost entirely limited to chronic stress models such as chronic social defeat or chronic mild stress, which are more technically challenging but show unique sensitivity to chronic and not acute antidepressant administration, comparable to the therapeutic delay of 4–6 weeks that is required for all available antidepressant drugs to treat depression in humans[2]. Animal models also have face validity, in which certain behavioural changes brought about by stress or genetic manipulation superficially resemble depressive symptoms. For instance, an animal's decreased sucrose intake after chronic stress is thought to model anhedonia[25].

These tests have been applied to the study of the molecular neurobiology of depression in three main ways. The most popular approach documents neuroplastic changes in brain regions after chronic stress and has revealed a role for structural, transcriptional and epigenetic changes in several brain regions (for example, those shown in Figs 1–3). These models can also be used to examine the behavioural effects of region-specific genetic manipulation, achieved through targeted genetic mutations in mice or virus-mediated gene transfer. The selective breeding of extreme populations within outbred rodents has also been used to generate stress-vulnerable or stress-resistant inbred strains (not shown)[17]. This approach is particularly valuable for quantitative trait locus (QTL) analyses, as well as for dissecting epigenetic contributions to responsiveness to stress[25]. These behavioural assays can also be used to study biological mechanisms that underlie phenotypic variations in stress responses. For example, susceptibility to social defeat is mediated by increases in the electrical activity of dopamine-producing neurons in the ventral tegmental area (VTA)[25], whereas resistance to learned helplessness is mediated by induction of the gene encoding the transcription factor ΔFOSB in the periaqueductal grey (PAG) area[84] (see box figure, right). In these ways, preclinical models of depression have provided important insights into the pathophysiology of depression.

largely on the role of brain-derived neurotrophic factor (BDNF), which is expressed abundantly in adult limbic structures. Support for this 'BDNF hypothesis' has come from a large preclinical literature showing that several forms of stress reduce BDNF-mediated signalling in the hippocampus, whereas chronic treatment with antidepressants increases BDNF-mediated signalling[2,31]. Similar changes have been observed in the post-mortem hippocampus of humans with depression[33], as well as in the concentrations of serum BDNF, the source of which remains controversial[31].

More causal evidence for the antidepressant action of BDNF has come from experiments in rodents in which antidepressant effects were observed on direct infusion of BDNF into the hippocampus[34] and were blocked on the conditional or inducible knockout of the gene encoding BDNF from forebrain regions[32,35]. However, more recent findings have necessitated a revision of this hypothesis. First, a substantial number of preclinical studies either have failed to show these patterns of changes induced by stress and by antidepressants, or have shown the opposite effects[36,37]. Second, male mice with conditional forebrain deletions of BDNF or its receptor do not show depression-like behaviour[35,38]. Third, in other regions — for example the VTA and NAc — BDNF exerts a potent pro-depressant effect: chronic stress increases the amount of BDNF within the NAc[39], and the direct infusion of BDNF into the VTA–NAc increases depression-related behaviours[25,40], whereas a selective knockout of the gene encoding BDNF from this circuit has antidepressant-like effects[39]. Finally, a single-nucleotide polymorphism (G196A; Val 66 → Met 66) in the gene encoding BDNF, which significantly impairs the intracellular trafficking and activity-dependent release of BDNF[41,42] and decreases hippocampal volume[41,43], does not alter genetic vulnerability to depression[8,44] (Fig. 2). In addition, recent studies suggest complex interactions between the *BDNF* G196A polymorphism, a polymorphism in the serotonin transporter gene, and stressful life events[45–47]. Taken together, these results suggest that the current formulation of the BDNF hypothesis is too simplistic; BDNF-mediated signalling is involved in neuroplastic responses to stress and antidepressants, but these effects are both region-specific[19] and antidepressant-specific[31] and function in the background of other potent genetic and environmental modifiers.

A marked cellular effect of several, but not all, antidepressant treatments is the induction of adult hippocampal neurogenesis — the process by which neural progenitors of the hippocampal subgranular zone (SGZ) divide mitotically to form new neurons that differentiate and integrate into the dentate gyrus[20,48]. Blockade of hippocampal neurogenesis inhibits the therapeutic-like effects of most antidepressant treatments in rodent models[48]. Moreover, treatment with antidepressants, possibly through the actions of CREB or other transcriptional

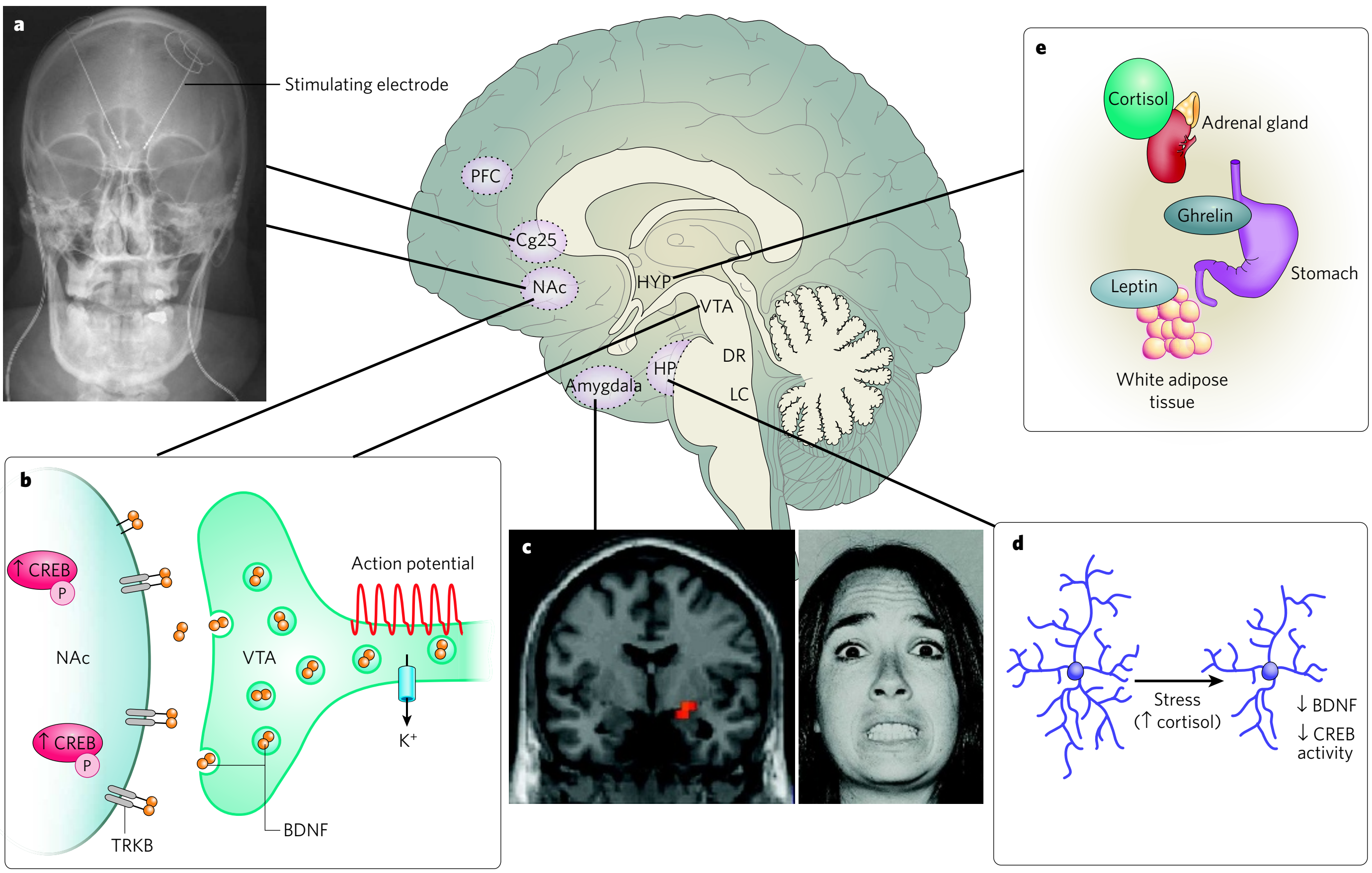


Figure 1 | Neural circuitry of depression. Several brain regions are implicated in the pathophysiology of depression. **a**, Deep brain stimulation of the subgenual cingulate cortex (Cg25)[17] or the nucleus accumbens (NAc)[18] has an antidepressant effect on individuals who have treatment-resistant depression. This effect is thought to be mediated through inhibiting the activity of these regions either by depolarization blockade or by stimulation of passing axonal fibres. (Image courtesy of T. Schlaepfer and V. Sturm, University Hospital, Bonn, Germany.) **b**, Increased activity-dependent release of brain-derived neurotrophic factor (BDNF) within the mesolimbic dopamine circuit (dopamine-producing ventral tegmental area (VTA) to dopamine-sensitive NAc) mediates susceptibility to social stress[25], probably occurring in part through activation of the transcription factor CREB (cyclic-AMP-response-element-binding protein)[20] by phosphorylation (P). **c**, Neuroimaging studies strongly implicate the amygdala (red pixels show activated areas) as an important limbic node for processing emotionally salient stimuli, such as fearful faces[7]. (Image courtesy of D. Weinberger, National Institute of Mental Health, Bethesda, Maryland). **d**, Stress decreases the concentrations of neurotrophins (such as BDNF), the extent of neurogenesis and the complexity of neuronal processes in the hippocampus (HP), effects that are mediated in part through increased cortisol concentrations and decreased CREB activity[2,14]. **e**, Peripherally released metabolic hormones in addition to cortisol, such as ghrelin[95] and leptin[96], produce mood-related changes through their effects on the hypothalamus (HYP) and several limbic regions (for example, the hippocampus, VTA and NAc). DR, dorsal raphe; LC, locus coeruleus; PFC, prefrontal cortex.

regulators[2,20], increases the amounts of several growth factors in the hippocampus that influence neurogenesis. These include BDNF (which promotes neuronal survival[49]), as well as vascular endothelial growth factor (VEGF) and VGF, which themselves have antidepressant and pro-neurogenic properties in rodents[50–52]. The mechanisms by which new neurons might restore mood are largely unknown. Activity-dependent increases in neurogenesis might increase activity propagation through hippocampal subfields[53] and allow hippocampal networks to adapt and learn new experiences[54]. Indeed, this raises the possibility that the presence of intact neurogenesis during stressful episodes mediates maladaptive learning and thus promotes depressive sequelae. Whereas several types of stress reduce SGZ cell proliferation, decreased neurogenesis does not itself produce depression[48,55]: rodents in which hippocampal neurogenesis has been ablated (through either irradiation[55,56] or genetic techniques[57]) do not show anxiety-related or depression-related behaviours.

Collectively, these studies highlight the weaknesses of attempts to generate a 'unified theory' of depression. Mechanisms that promote depressive symptoms in response to stress differ markedly between different neural circuits and can also be distinct from changes that underlie depression in the absence of external stress ('endogenous depression'). In addition, neuroplastic events that are required for antidepressant efficacy need not function through the reversal of stress-induced plasticity[2], and might function through separate and parallel circuits.

Neuroendocrine and neuroimmune interactions

Early clinical studies identifying reproducible but small increases in serum glucocorticoid concentrations in depression[58,59] fuelled significant interest in the role of a dysfunctional hypothalamic–pituitary–adrenal axis in the pathophysiology of depression. Physical or psychological stress increases serum glucocorticoid concentrations, and some depression-like symptoms can be produced in rodents by chronic administration of glucocorticoids[60]. Excess glucocorticoids, through the activation of glucocorticoid receptors, can reduce SGZ proliferation rates and produce atrophic changes in hippocampal subregions[61]. This could contribute to the hippocampal volume reductions seen in depression. Patients with Cushing's syndrome, who have extremely high concentrations of circulating cortisol, also show depressive features and atrophic changes in the hippocampus[2,61]. Several metabolic abnormalities that are often associated with depression, such as insulin resistance and abdominal obesity, can be at least partly explained by an increase in glucocorticoids[4,62]. Hypercortisolaemia in depression is manifested at several levels, including impaired glucocorticoid-receptor-mediated negative feedback[62], adrenal hyper-responsiveness to circulating adrenocorticotropic hormone (ACTH)[58] and hypersecretion of corticotropin-releasing factor[63], the hypothalamic activator of ACTH release from the pituitary[2,64]. In line with these findings, glucocorticoid and corticotropin-releasing factor receptor antagonists are currently being tested in clinical trials[27].

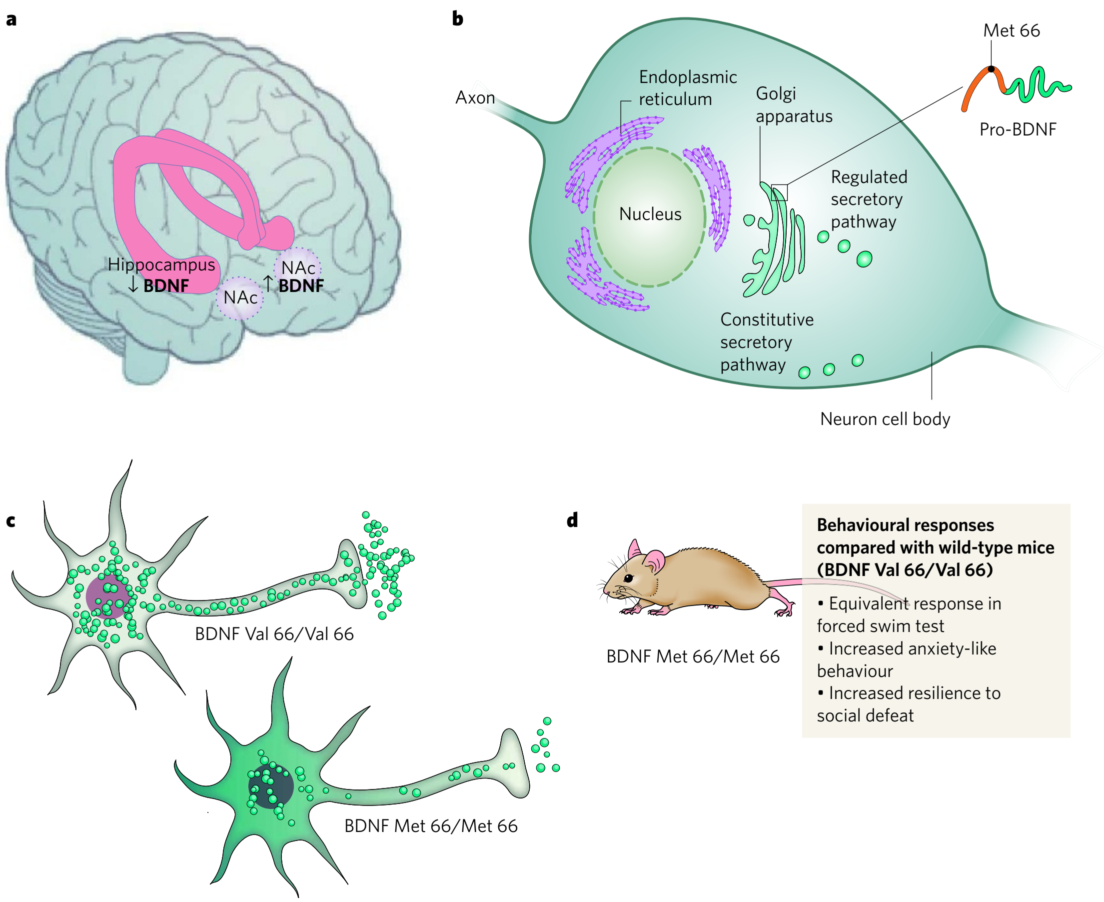


Figure 2 | BDNF and depression — an example of the complexities of the molecular pathophysiology of depression. **a**, Post-mortem data from depressed humans show that depression is associated with a decrease in the amount of BDNF in the hippocampus[33] and an increase (of similar magnitude) in the NAc[25], an example of the regional specificity of depression-related neuroplastic changes. **b**, Neuronal secretion of BDNF occurs through regulated (activity-dependent) and constitutive secretory pathways. Regulated secretion is modulated by the interactions of proteins in the Golgi apparatus with the pro-domain of BDNF, the site of a single-nucleotide polymorphism (G196A) in humans that results in the substitution of valine at amino-acid residue 66 with methionine. **c**, The Met-66-containing BDNF variant has impaired intracellular trafficking. Met-66 BDNF is not properly sorted within the cell, causing it to be distributed throughout the cell body outside of vesicles[42]. In addition, less BDNF is secreted from the nerve terminal. **d**, Knock-in mice that homozygously express Met-66 BDNF[41] have normal responses in the forced-swim test[25], but these mice show more anxiety-like behaviour[41] and greater resilience to behavioural and molecular changes after social defeat[25], implicating this BDNF polymorphism in the pathophysiology of psychological disorders that are influenced by stressful life events.

More recent studies suggest that hypercortisolaemia is almost exclusively a feature of very severe depressive episodes, such as are observed in an in-patient setting[65] or accompanied by psychotic symptoms (for example, hallucinations and delusions)[2,9] in which glucocorticoid antagonists show some therapeutic efficacy[66]. By contrast, atypical depression, a subtype characterized by hyperphagia and hypersomnia, seems to be associated with hypocortisolaemia[65,67], a phenomenon that is also observed in certain associated conditions such as fibromyalgia, chronic fatigue syndrome and post-traumatic stress disorder[68]. The origins of such distinct glucocorticoid profiles might reflect the evolutionary trade-off between the catabolic and immunosuppressant effects of glucocorticoids: whereas high serum concentrations of glucocorticoids promote the mobilization of energy resources during stressful experiences, low glucocorticoid states allow an unobstructed immune system to combat infection or physical injury sustained during adverse encounters in the wild[59].

Cytokines, which are humoral mediators of innate and adaptive immunity, are also important modulators of mood. Cytokine receptors within the central nervous system are activated by both peripherally and centrally synthesized cytokines[69]. Low doses of lipopolysaccharide or interleukin 1 (IL-1) produce 'sickness behaviour' in rodents (consisting of social withdrawal and decreased exploratory and sexual behaviour), brought about by the release of pro-inflammatory cytokines such as interferon-α, tumour necrosis factor-α (TNF-α), IL-6 and IL-1β, which activate the hypothalamic–pituitary–adrenal axis and central monoamine systems[70]. Roughly 30% of individuals treated with recombinant interferons develop depression as a side effect of treatment[71]. Clinical studies examining depression-associated increases in serum cytokine concentrations have been largely inconsistent[70]. This suggests that immune activation is a signature of a small subset of depression cases, including those associated with autoimmune conditions such as rheumatoid arthritis, in which heightened system-wide inflammation can increase the risk of acute coronary events[4] in addition to producing depressive mood changes.

Administration of cytokines such as interferon-α or IL-6 to rodents does not cause consistent depression-like features[70]. Nevertheless, recent preclinical studies indicate that blocking pro-inflammatory cytokine-mediated signalling can produce antidepressant effects. Mice with targeted deletions of the gene encoding IL-6 (ref. 72) or those encoding the TNF-α receptors[73] show antidepressant-like behavioural phenotypes, and a centrally administered antagonist of the IL-1β receptor reversed the behavioural and antineurogenic effects of chronic stress[74]. Future studies of the 'cytokine hypothesis' must focus on elucidating the largely unknown neural circuitry involved in the behavioural effects of cytokines, and must more precisely delineate the intercellular interactions involved between brain macrophages (microglia), glia and neurons within this circuitry.

Epigenetic mechanisms

Among the several methods by which experience can produce long-lasting changes in protein availability and function, there has been considerable recent interest in epigenetic modifications in the pathophysiology of depression and antidepressant action. These modifications (Fig. 3) encompass covalent changes to DNA (for example, DNA methylation) and post-translational modifications of histone N-terminal tails (for example, acetylation and methylation), as well as non-transcriptional gene-silencing mechanisms (for example, microRNAs)[75]. Given that these changes can be long-lasting, epigenetics has been invoked to explain several aspects of depression, including high discordance rates between monozygotic twins, individual differences among inbred rodents, the chronic relapsing nature of the illness, and the strikingly greater prevalence of depression in women[11]. In essence, epigenetic changes offer a mechanism by which environmental experiences can modify gene function in the absence of DNA sequence changes, and they might help to explain largely inconsistent genetic association studies of depression, for example by undermining the transcriptional impact of DNA sequence polymorphisms due to epigenetic modifications on those gene promoters[11]. Although epigenetic changes have been implicated in numerous psychiatric conditions[75], the field of depression research has focused on two main chromatin-modifying processes. The first is DNA methylation (of cytosine), which seems to be important in the influence of maternal behaviour on adult emotional processing. Adult offspring of rats born to mothers with low rates of maternal licking and grooming show increased anxiety and reduced expression of glucocorticoid receptors within the hippocampus compared with offspring of mothers with high rates of maternal behaviours. This reduced expression of glucocorticoid receptors is mediated by increased methylation of the glucocorticoid receptor gene promoter (effectively repressing gene expression). This long-lasting 'molecular scar'[75] is established within the first week of life and is effectively reversed by cross-fostering[76]. Interestingly, this increase in methylation was also reversed by the infusion of trichostatin A, a histone deacetylase (HDAC) inhibitor[77].

Histone acetylation, which is associated with transcriptional activation and decondensed chromatin, seems to be a key substrate for antidepressant action[78]. Increased histone acetylation at the *Bdnf* promoter in the hippocampus was shown to be required for the ability of chronically administered imipramine to reverse certain deleterious effects of social defeat[79]. Moreover, HDAC inhibitors show antidepressant-like effects in the social-defeat assay and other behavioural assays[79,80], and efforts are underway to develop more potent agents that are designed to target specific HDACs, such as HDAC5, a class II HDAC[75,79]. The implications of these studies come with an important anatomical caveat: although inhibiting the actions of HDAC5 in the hippocampus seems to be therapeutically advantageous[17,80], mice that are globally deficient in HDAC5 are more vulnerable to social defeat[81]. Similarly, although imipramine increases HDAC5 expression in the hippocampus[79], it significantly reduces HDAC5 expression within the NAc[81], further emphasizing the regional specificity of stress-related and antidepressant-related plasticity.

Current knowledge of the diversity of chromatin-modifying enzymes, and techniques to detect and quantify chromatin modifications genome-wide, is growing at an enormous pace. An important challenge in the clinical translation of these approaches will be to improve the technological ability to demonstrate causation by developing techniques to detect these modifications *in vivo*. Such techniques will allow researchers to examine, for the first time, region-specific chromatin measures associated with depression or antidepressant responses in humans.

Resilience-related research

Humans show a remarkable heterogeneity in their responses to stress and adversity: although a subset of depression cases can be causally attributed to stressful life events, these events in themselves raise only moderately the risk of developing depression[10]. In addition, reactive dysphoric states such as post-traumatic stress disorder only emerge in about 10–20% of trauma-exposed individuals[82]. Although a large body of research describes maladaptive neurobiological changes that occur after stressful exposures (such as decreased hippocampal neurogenesis and lower concentrations of BDNF, as discussed in the section 'Neurotrophins and neurogenesis'), relatively little attention has been devoted to understanding how most individuals adapt well — that is, are 'resilient' — in the face of adversity[83].

Animal models have recently been used to provide some neurobiological insight into these clinical observations. For example, by exploiting natural variations in the development of active escape in the learned-helplessness test, stress-induced upregulation of the transcription factor ΔFOSB (a stable, truncated protein product of the *Fosb* gene) in the midbrain periaqueductal grey nucleus was shown to promote a resilient phenotype. This effect was mediated through downregulating expression of substance P, a neuropeptide released during stress[84]. A more recent report illustrated the role of mesolimbic dopamine-mediated signalling in emotional homeostatic mechanisms[25]. By adapting the social-defeat model[39,79] of depression to examine the variations in response to chronic stress[85], vulnerability to the development of social avoidance and other

deleterious sequelae was shown to be mediated by the increased excitability of VTA dopamine neurons and their subsequent increased activity-dependent release of BDNF onto NAc neurons. Resilient mice (which also show increased ΔFOSB concentrations[84]) escaped this increase in VTA neuronal excitability by upregulating voltage-gated potassium channels, which functions as a molecular compensation to restore normal excitability and maintain low levels of BDNF-mediated signalling in the NAc. Other putative mechanisms of resilience have come from clinical and preclinical investigations. One involves the release of neuropeptide Y from locus coeruleus nerve terminals onto amygdala neurons, which promotes resilient behaviour[83,86]. Interestingly, many of these studies report stable individual differences in stress responses among genetically inbred mice, strongly implicating non-genomic factors[25,84,87]. As these mice are housed under identical environmental conditions as well, the findings indicate the likely importance of epigenetic mechanisms during development, a possibility that now requires direct investigation.

Gene expression profiling of stress-vulnerable and stress-resilient mice revealed distinct transcriptional profiles in the VTA and NAc[19], and similar results have been obtained in the hippocampus with related methods[88]. These findings suggest that resilient behaviour represents a distinct, active neurobiological process (not simply the absence of vulnerability)[25]. Accordingly, a comprehensive understanding of such molecular mechanisms of allostasis (ongoing efforts to maintain homeostasis)[61] has the potential to be harnessed for the development of new therapeutic agents. In these ways, the identification of antivulnerability processes will be an important alternative approach to improving knowledge about the neurobiology of stress and the pathophysiology of depression.

New insights

Although the hypotheses described here remain active areas of research, recent findings have sparked interest in neurobiological systems that were previously unexplored in depression. A dramatic example is the observation that sub-anaesthetic doses of intravenously infused ketamine (a non-competitive NMDA (*N*-methyl-D-aspartate) glutamate receptor antagonist and psychotomimetic) produce a rapid but transient antidepressant effect on individuals with treatment-resistant depression[89]. This effect suggests that depressive symptoms can be improved by altering the actions of glutamate, the major excitatory neurotransmitter in the brain. The antidepressant properties of ketamine have been recapitulated in animal tests of antidepressant action, such as the forced-swim test, in which the ability of ketamine to reduce immobility required intact signalling through AMPA receptors for glutamate[90] and was associated with increased concentrations of hippocampal BDNF protein[91]. Despite the limited evidence for dysfunction in specific glutamate systems in depression, the clinical effects of ketamine have inspired new lines of preclinical research to explore the underlying neural circuitry and downstream signalling, as well as to identify previously unidentified NMDA receptor modulators that could be targeted to achieve better side-effect profiles[92].

In the past few years, there has also been an increased interest in examining interactions between traditional mood substrates and pathways involved in the control of feeding and metabolism[29]. MCH (melanin-concentrating hormone)-containing neurons projecting from the lateral hypothalamus to several limbic regions including NAc provide an important orexigenic (pro-appetite) signal. Global decreases in MCH-mediated signalling[93], as well as local MCH antagonism within the NAc[94], produce antidepressant-like responses in several rodent models, generating tremendous interest in the antidepressant potential of selective MCH antagonists[14], which might also curb the weight gain associated with a subset of depression[19]. In contrast to the pro-depressant actions of MCH, other peptides such as orexin and ghrelin might have an antidepressant role, particularly during conditions of low caloric intake[95]. These and other studies illustrate the general theme that an animal's metabolic status greatly influences mood and motivation. Understanding the complex molecular interactions between peripheral metabolic signals (such as ghrelin[95] and leptin[96]) and centrally released regulators of feeding and arousal (such as MCH,

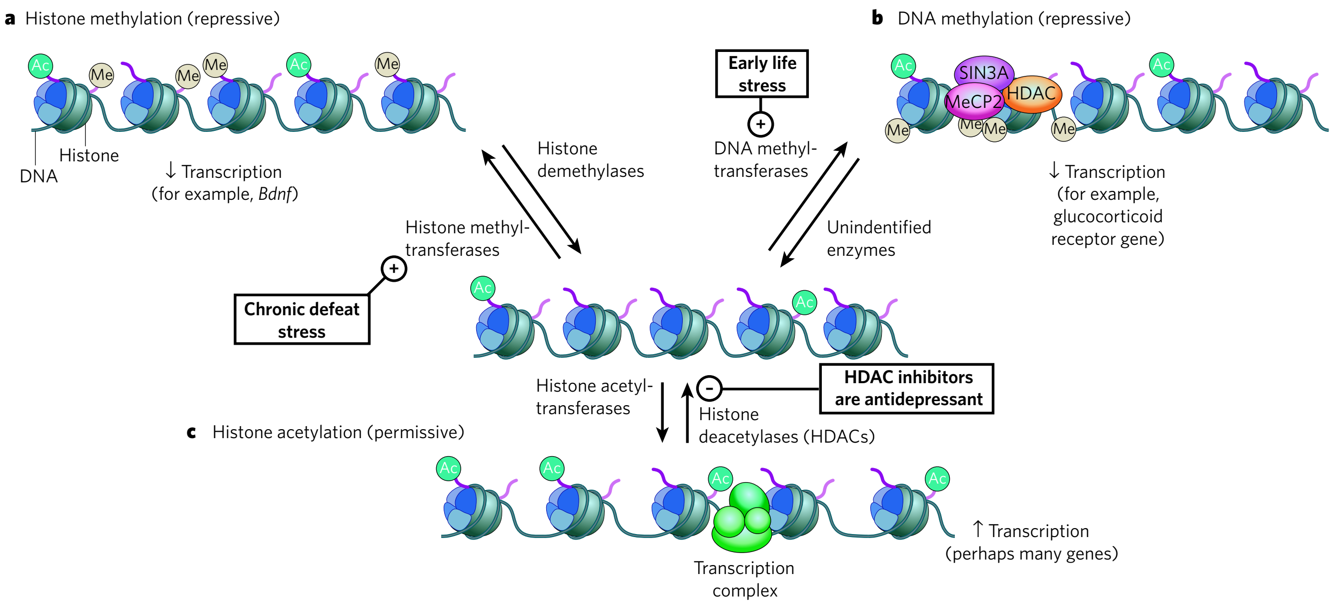


Figure 3 | Epigenetic regulation in depression. The transcriptional potential of genes involved in neuroplastic responses to stress or antidepressant treatments can be regulated through chromatin-remodelling events catalysed by specific enzymes. **a**, The methylation of histones on specific lysine residues (for example, Lys 9 and Lys 27) is associated with condensed chromatin (heterochromatin) and is important in the repression of *Bdnf* expression in the hippocampus after social defeat[79]. The pluses and minuses indicate activation or inhibition, respectively, of a particular process. **b**, By contrast, repression of other genes can occur through the methylation of cytosine within CpG islands of promoter regions, attracting proteins involved in transcriptional repression, such as SIN3A, MeCP2 (methyl-CpG binding protein 2) and histone deacetylases (HDACs). DNA methylation of the promoter of the glucocorticoid receptor gene occurs in rat pups born to mothers with inherently low levels of maternal behaviour[77]. Although such methylation events have been reported to be reversible, the enzymes responsible for demethylating DNA have yet to be identified[75,76]. **c**, Histone acetylation, catalysed by histone acetyltransferases, is associated with decondensed chromatin (euchromatin), increasing the activity of transcriptional complexes. HDAC inhibitors (which activate the expression of numerous genes that have not yet been identified with certainty) show antidepressant properties in several assays[79,80]. Ac, acetyl; Me, methyl.

orexin, neuropeptide Y[83] and α-melanocyte-stimulating hormone[97]) might provide new pathophysiological and therapeutic insights into mood disorders.

Conclusion

Knowledge of the pathophysiology of depression has evolved substantially: from Galen's speculations in antiquity about an excess of black bile ('*melancholia*')[2,9] to theories focused on 'psychic pain' and 'chemical imbalances', and then to more current hypotheses that incorporate gene–environment interactions, endocrine, immunological and metabolic mediators, and cellular, molecular and epigenetic forms of plasticity. However, enormous gaps in the knowledge of depression and its treatment persist. Instead of being overwhelmed by the heterogeneity of the illness, researchers and clinicians must embrace the polysyndromic nature of depression and use a multidisciplinary approach to explore the neurobiological bases for depression's many subtypes. To improve the still-low remission rates[21], it will be imperative to look beyond monoamine and neurotrophic mechanisms[14] and expand knowledge about antidepressant pharmacogenetics. Researchers must better understand the biological basis for the efficacy of deep brain stimulation in depression, and must explore the therapeutic possibilities of viral-mediated gene delivery, which is being applied successfully to other neuropsychiatric disorders[98]. Finally, the field must harness the full potential of preclinical studies by continuing to develop improved animal models that incorporate the powerful array of molecular and anatomical tools available today, and must follow a systems approach to the study of depression that acknowledges the powerful bidirectional interactions between peripheral organs and the brain. ■

1. Kessler, R. C. *et al.* Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch. Gen. Psychiatry* **62,** 593–602 (2005).
2. Nestler, E. J. *et al.* Neurobiology of depression. *Neuron* **34,** 13–25 (2002).
3. Knol, M. J. *et al.* Depression as a risk factor for the onset of type 2 diabetes mellitus. A meta-analysis. *Diabetologia* **49,** 837–845 (2006).
4. Evans, D. L. *et al.* Mood disorders in the medically ill: scientific review and recommendations. *Biol. Psychiatry* **58,** 175–189 (2005).
5. Gildengers, A. G. *et al.* Medical burden in late-life bipolar and major depressive disorders. *Am. J. Geriatr. Psychiatry* **16,** 194–200 (2008).
6. Phelps, E. A. & LeDoux, J. E. Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron* **48,** 175–187 (2005).
7. Drevets, W. C. Neuroimaging and neuropathological studies of depression: implications for the cognitive-emotional features of mood disorders. *Curr. Opin. Neurobiol.* **11,** 240–249 (2001).
8. Lopez-Leon, S. *et al.* Meta-analyses of genetic studies on major depressive disorder. *Mol. Psychiatry* **13,** 772–785 (2007).
9. Rush, A. J. The varied clinical presentations of major depressive disorder. *J. Clin. Psychiatry* **68** (Suppl. 8), 4–10 (2007).
10. Kendler, K. S., Karkowski, L. M. & Prescott, C. A. Causal relationship between stressful life events and the onset of major depression. *Am. J. Psychiatry* **156,** 837–841 (1999).
11. Mill, J. & Petronis, A. Molecular studies of major depressive disorder: the epigenetic perspective. *Mol. Psychiatry* **12,** 799–814 (2007).
12. Hasler, G., Drevets, W. C., Manji, H. K. & Charney, D. S. Discovering endophenotypes for major depression. *Neuropsychopharmacology* **29,** 1765–1781 (2004).
13. Ressler, K. J. & Mayberg, H. S. Targeting abnormal neural circuits in mood and anxiety disorders: from the laboratory to the clinic. *Nature Neurosci.* **10,** 1116–1124 (2007).
14. Berton, O. & Nestler, E. J. New approaches to antidepressant drug discovery: beyond monoamines. *Nature Rev. Neurosci.* **7,** 137–151 (2006).
15. Sheline, Y. I. Neuroimaging studies of mood disorder effects on the brain. *Biol. Psychiatry* **54,** 338–352 (2003).
16. Harrison, P. J. The neuropathology of primary mood disorder. *Brain* **125,** 1428–1449 (2002).
17. Mayberg, H. S. *et al.* Deep brain stimulation for treatment-resistant depression. *Neuron* **45,** 651–660 (2005).
This paper gives the first demonstration of the therapeutic efficacy of deep brain stimulation applied to the subgenual cingulate gyrus for treatment-refractory depression.
18. Schlaepfer, T. E. *et al.* Deep brain stimulation to reward circuitry alleviates anhedonia in refractory major depression. *Neuropsychopharmacology* **33,** 368–377 (2008).
19. Nestler, E. J. & Carlezon, W. A. Jr. The mesolimbic dopamine reward circuit in depression. *Biol. Psychiatry* **59,** 1151–1159 (2006).
20. Pittenger, C. & Duman, R. S. Stress, depression, and neuroplasticity: a convergence of mechanisms. *Neuropsychopharmacology* **33,** 88–109 (2008).
21. Trivedi, M. H. *et al.* Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am. J. Psychiatry* **163,** 28–40 (2006).
22. Ansorge, M. S., Hen, R. & Gingrich, J. A. Neurodevelopmental origins of depressive disorders. *Curr. Opin. Pharmacol.* **7,** 8–17 (2007).
23. Ruhe, H. G., Mason, N. S. & Schene, A. H. Mood is indirectly related to serotonin, norepinephrine and dopamine levels in humans: a meta-analysis of monoamine depletion studies. *Mol. Psychiatry* **12,** 331–359 (2007).
24. Hu, H. *et al.* Emotion enhances learning via norepinephrine regulation of AMPA-receptor trafficking. *Cell* **131,** 160–173 (2007).
25. Krishnan, V. *et al.* Molecular adaptations underlying susceptibility and resistance to social defeat in brain reward regions. *Cell* **131,** 391–404 (2007).
This paper adapts the social-defeat model to study resilient behaviour and identifies active neurobiological mechanisms that maintain normal functioning in the face of chronic stress.
26. Svenningsson, P. *et al.* Alterations in 5-HT1B receptor function by p11 in depression-like states. *Science* **311,** 77–80 (2006).
27. Mathew, S. J., Manji, H. K. & Charney, D. S. Novel drugs and therapeutic targets for severe mood disorders. *Neuropsychopharmacology* **33,** 2080–2092 (2008).
This paper provides an up-to-date and comprehensive list of new antidepressant drugs currently in various stages of clinical trials.
28. Lucas, G. *et al.* Serotonin$_4$ (5-HT$_4$) receptor agonists are putative antidepressants with a rapid onset of action. *Neuron* **55,** 712–725 (2007).
29. Uhr, M. *et al.* Polymorphisms in the drug transporter gene *ABCB1* predict antidepressant treatment response in depression. *Neuron* **57,** 203–209 (2008).
30. Holsboer, F. How can we realize the promise of personalized antidepressant medicines? *Nature Rev. Neurosci.* **9,** 638–646 (2008).
31. Duman, R. S. & Monteggia, L. M. A neurotrophic model for stress-related mood disorders. *Biol. Psychiatry* **59,** 1116–1127 (2006).
32. Monteggia, L. M. *et al.* Essential role of brain-derived neurotrophic factor in adult hippocampal function. *Proc. Natl Acad. Sci. USA* **101,** 10827–10832 (2004).
33. Karege, F., Vaudan, G., Schwald, M., Perroud, N. & La Harpe, R. Neurotrophin levels in postmortem brains of suicide victims and the effects of antemortem diagnosis and psychotropic drugs. *Brain Res. Mol. Brain Res.* **136,** 29–37 (2005).
34. Shirayama, Y., Chen, A. C., Nakagawa, S., Russell, D. S. & Duman, R. S. Brain-derived neurotrophic factor produces antidepressant effects in behavioral models of depression. *J. Neurosci.* **22,** 3251–3261 (2002).
35. Monteggia, L. M. *et al.* Brain-derived neurotrophic factor conditional knockouts show gender differences in depression-related behaviors. *Biol. Psychiatry* **61,** 187–197 (2007).
36. Groves, J. O. Is it time to reassess the BDNF hypothesis of depression? *Mol. Psychiatry* **12,** 1079–1088 (2007).
37. Martinowich, K., Manji, H. & Lu, B. New insights into BDNF function in depression and anxiety. *Nature Neurosci.* **10,** 1089–1093 (2007).
38. Zorner, B. *et al.* Forebrain-specific trkB-receptor knockout mice: behaviorally more hyperactive than 'depressive'. *Biol. Psychiatry* **54,** 972–982 (2003).
39. Berton, O. *et al.* Essential role of BDNF in the mesolimbic dopamine pathway in social defeat stress. *Science* **311,** 864–868 (2006).
By using the social-defeat model, this paper characterizes the potent, pro-depressant effects of BDNF in the mesolimbic dopamine pathway, which is opposite to the well-described antidepressant-like effects of BDNF in hippocampal circuits.
40. Eisch, A. J. *et al.* Brain-derived neurotrophic factor in the ventral midbrain-nucleus accumbens pathway: a role in depression. *Biol. Psychiatry* **54,** 994–1005 (2003).
41. Chen, Z. Y. *et al.* Genetic variant BDNF (Val66Met) polymorphism alters anxiety-related behavior. *Science* **314,** 140–143 (2006).
42. Egan, M. F. *et al.* The BDNF Val66Met polymorphism affects activity-dependent secretion of BDNF and human memory and hippocampal function. *Cell* **112,** 257–269 (2003).
This multidisciplinary study demonstrates how the polymorphism in BDNF in which methionine is substituted for valine at position 66 causes deficits in episodic memory, alters hippocampal activation and decreases activity-dependent BDNF secretion.
43. Szeszko, P. R. *et al.* Brain-derived neurotrophic factor Val66Met polymorphism and volume of the hippocampal formation. *Mol. Psychiatry* **10,** 631–636 (2005).
44. Gratacos, M. *et al.* Brain-derived neurotrophic factor Val66Met and psychiatric disorders: meta-analysis of case-control studies confirm association to substance-related disorders, eating disorders, and schizophrenia. *Biol. Psychiatry* **61,** 911–922 (2007).
45. Kaufman, J. *et al.* Brain-derived neurotrophic factor-5-HTTLPR gene interactions and environmental modifiers of depression in children. *Biol. Psychiatry* **59,** 673–680 (2006).
46. Kim, J. M. *et al.* Interactions between life stressors and susceptibility genes (5-HTTLPR and BDNF) on depression in Korean elders. *Biol. Psychiatry* **62,** 423–428 (2007).
47. Pezawas, L. *et al.* Evidence of biologic epistasis between BDNF and SLC6A4 and implications for depression. *Mol. Psychiatry* **13,** 709–716 (2008).
48. Sahay, A. & Hen, R. Adult hippocampal neurogenesis in depression. *Nature Neurosci.* **10,** 1110–1115 (2007).
49. Sairanen, M., Lucas, G., Ernfors, P., Castren, M. & Castren, E. Brain-derived neurotrophic factor and antidepressant drugs have different but coordinated effects on neuronal turnover, proliferation, and survival in the adult dentate gyrus. *J. Neurosci.* **25,** 1089–1094 (2005).
50. Hunsberger, J. G. *et al.* Antidepressant actions of the exercise-regulated gene VGF. *Nature Med.* **13,** 1476–1482 (2007).
This paper characterizes the antidepressant effects of VGF, an exercise-regulated neurotrophic factor in the hippocampus, and identifies VGF-mediated signalling as a potential therapeutic target.
51. Thakker-Varia, S. *et al.* The neuropeptide VGF produces antidepressant-like behavioral effects and enhances proliferation in the hippocampus. *J. Neurosci.* **27,** 12156–12167 (2007).
52. Warner-Schmidt, J. L. & Duman, R. S. VEGF is an essential mediator of the neurogenic and behavioral actions of antidepressants. *Proc. Natl Acad. Sci. USA* **104,** 4647–4652 (2007).
53. Airan, R. D. *et al.* High-speed imaging reveals neurophysiological links to behavior in an animal model of depression. *Science* **317,** 819–823 (2007).
This study uses quantitative voltage-sensitive dye imaging to explore the contribution of antidepressant-induced neurogenesis to local hippocampal network dynamics.
54. Kempermann, G. The neurogenic reserve hypothesis: what is adult hippocampal neurogenesis good for? *Trends Neurosci.* **31,** 163–169 (2008).
55. Surget, A. *et al.* Drug-dependent requirement of hippocampal neurogenesis in a model of depression and of antidepressant reversal. *Biol. Psychiatry* **64,** 293–301 (2008).
56. Santarelli, L. *et al.* Requirement of hippocampal neurogenesis for the behavioral effects of antidepressants. *Science* **301,** 805–809 (2003).
57. Zhao, C., Deng, W. & Gage, F. H. Mechanisms and functional implications of adult neurogenesis. *Cell* **132,** 645–660 (2008).

58. Parker, K. J., Schatzberg, A. F. & Lyons, D. M. Neuroendocrine aspects of hypercortisolism in major depression. *Horm. Behav.* **43,** 60–66 (2003).
59. Raison, C. L. & Miller, A. H. When not enough is too much: the role of insufficient glucocorticoid signaling in the pathophysiology of stress-related disorders. *Am. J. Psychiatry* **160,** 1554–1565 (2003).
60. Gourley, S. L. *et al.* Regionally specific regulation of ERK MAP kinase in a model of antidepressant-sensitive chronic depression. *Biol. Psychiatry* **63,** 353–359 (2007).
61. McEwen, B. S. Physiology and neurobiology of stress and adaptation: central role of the brain. *Physiol. Rev.* **87,** 873–904 (2007).
62. Brown, E. S., Varghese, F. P. & McEwen, B. S. Association of depression with medical illness: does cortisol play a role? *Biol. Psychiatry* **55,** 1–9 (2004).
63. Nemeroff, C. B. & Owens, M. J. Treatment of mood disorders. *Nature Neurosci.* **5** (Suppl.), 1068–1070 (2002).
64. de Kloet, E. R., Joels, M. & Holsboer, F. Stress and the brain: from adaptation to disease. *Nature Rev. Neurosci.* **6,** 463–475 (2005).
65. Brouwer, J. P. *et al.* Thyroid and adrenal axis in major depression: a controlled study in outpatients. *Eur. J. Endocrinol.* **152,** 185–191 (2005).
66. Schatzberg, A. F. & Lindley, S. Glucocorticoid antagonists in neuropsychotic disorders. *Eur. J. Pharmacol.* **583,** 358–364 (2008).
67. Gold, P. W. & Chrousos, G. P. Organization of the stress system and its dysregulation in melancholic and atypical depression: high vs low CRH/NE states. *Mol. Psychiatry* **7,** 254–275 (2002).
68. Heim, C., Ehlert, U. & Hellhammer, D. H. The potential role of hypocortisolism in the pathophysiology of stress-related bodily disorders. *Psychoneuroendocrinology* **25,** 1–35 (2000).
69. Dantzer, R., O'Connor, J. C., Freund, G. G., Johnson, R. W. & Kelley, K. W. From inflammation to sickness and depression: when the immune system subjugates the brain. *Nature Rev. Neurosci.* **9,** 46–56 (2008).
70. Dunn, A. J., Swiergiel, A. H. & de Beaurepaire, R. Cytokines as mediators of depression: what can we learn from animal studies? *Neurosci. Biobehav. Rev.* **29,** 891–909 (2005).
71. Loftis, J. M. & Hauser, P. The phenomenology and treatment of interferon-induced depression. *J. Affect. Disord.* **82,** 175–190 (2004).
72. Chourbaji, S. *et al.* IL-6 knockout mice exhibit resistance to stress-induced development of depression-like behaviors. *Neurobiol. Dis.* **23,** 587–594 (2006).
73. Simen, B. B., Duman, C. H., Simen, A. A. & Duman, R. S. TNFα signaling in depression and anxiety: behavioral consequences of individual receptor targeting. *Biol. Psychiatry* **59,** 775–785 (2006).
74. Koo, J. W. & Duman, R. S. IL-1β is an essential mediator of the antineurogenic and anhedonic effects of stress. *Proc. Natl Acad. Sci. USA* **105,** 751–756 (2008).
75. Tsankova, N., Renthal, W., Kumar, A. & Nestler, E. J. Epigenetic regulation in psychiatric disorders. *Nature Rev. Neurosci.* **8,** 355–367 (2007).
76. Szyf, M., Weaver, I. & Meaney, M. Maternal care, the epigenome and phenotypic differences in behavior. *Reprod. Toxicol.* **24,** 9–19 (2007).
77. Weaver, I. C. *et al.* Epigenetic programming by maternal behavior. *Nature Neurosci.* **7,** 847–854 (2004).
78. Tsankova, N. M., Kumar, A. & Nestler, E. J. Histone modifications at gene promoter regions in rat hippocampus after acute and chronic electroconvulsive seizures. *J. Neurosci.* **24,** 5603–5610 (2004).
79. Tsankova, N. M. *et al.* Sustained hippocampal chromatin regulation in a mouse model of depression and antidepressant action. *Nature Neurosci.* **9,** 519–525 (2006).
This paper contains one of the first demonstrations of the role of epigenetic modifications in stress-induced effects on the brain and their reversal by antidepressant treatments, as well as the therapeutic potential of HDAC inhibitors in depression.
80. Schroeder, F. A., Lin, C. L., Crusio, W. E. & Akbarian, S. Antidepressant-like effects of the histone deacetylase inhibitor, sodium butyrate, in the mouse. *Biol. Psychiatry* **62,** 55–64 (2007).
81. Renthal, W. *et al.* Histone deacetylase 5 epigenetically controls behavioral adaptations to chronic emotional stimuli. *Neuron* **56,** 517–529 (2007).
82. Yehuda, R. Risk and resilience in posttraumatic stress disorder. *J. Clin. Psychiatry* **65** (Suppl. 1)**,** 29–36 (2004).
83. Charney, D. S. Psychobiological mechanisms of resilience and vulnerability: implications for successful adaptation to extreme stress. *Am. J. Psychiatry* **161,** 195–216 (2004).
84. Berton, O. *et al.* Induction of ΔFosB in the periaqueductal gray by stress promotes active coping responses. *Neuron* **55,** 289–300 (2007).
In this paper, the learned-helplessness model is used to illustrate the pro-resilient effects of the transcription factor ΔFOSB within the dorsal raphe nucleus.
85. Strekalova, T., Spanagel, R., Bartsch, D., Henn, F. A. & Gass, P. Stress-induced anhedonia in mice is associated with deficits in forced swimming and exploration. *Neuropsychopharmacology* **29,** 2007–2017 (2004).
86. Sajdyk, T. J. *et al.* Neuropeptide Y in the amygdala induces long-term resilience to stress-induced reductions in social responses but not hypothalamic–adrenal–pituitary axis activity or hyperthermia. *J. Neurosci.* **28,** 893–903 (2008).
87. Peaston, A. E. & Whitelaw, E. Epigenetics and phenotypic variation in mammals. *Mamm. Genome* **17,** 365–374 (2006).
88. Bergstrom, A., Jayatissa, M. N., Thykjaer, T. & Wiborg, O. Molecular pathways associated with stress resilience and drug resistance in the chronic mild stress rat model of depression: a gene expression study. *J. Mol. Neurosci.* **33,** 201–215 (2007).
89. Zarate, C. A. Jr *et al.* A randomized trial of an N-methyl-D-aspartate antagonist in treatment-resistant major depression. *Arch. Gen. Psychiatry* **63,** 856–864 (2006).
90. Maeng, S. *et al.* Cellular mechanisms underlying the antidepressant effects of ketamine: role of α-amino-3-hydroxy-5-methylisoxazole-4-propionic acid receptors. *Biol. Psychiatry* **63,** 349–352 (2008).
91. Garcia, L. S. *et al.* Acute administration of ketamine induces antidepressant-like effects in the forced swimming test and increases BDNF levels in the rat hippocampus. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **32,** 140–144 (2008).
92. Maeng, S. & Zarate, C. A. Jr. The role of glutamate in mood disorders: results from the ketamine in major depression study and the presumed cellular mechanism underlying its antidepressant effects. *Curr. Psychiatry Rep.* **9,** 467–474 (2007).
93. Roy, M., David, N., Cueva, M. & Giorgetti, M. A study of the involvement of melanin-concentrating hormone receptor 1 (MCHR1) in murine models of depression. *Biol. Psychiatry* **61,** 174–180 (2007).
94. Georgescu, D. *et al.* The hypothalamic neuropeptide melanin-concentrating hormone acts in the nucleus accumbens to modulate feeding behavior and forced-swim performance. *J. Neurosci.* **25,** 2933–2940 (2005).
95. Lutter, M. *et al.* The orexigenic hormone ghrelin defends against depressive symptoms of chronic stress. *Nature Neurosci.* **11,** 752–753 (2008).
96. Lu, X. Y., Kim, C. S., Frazer, A. & Zhang, W. Leptin: a potential novel antidepressant. *Proc. Natl Acad. Sci. USA* **103,** 1593–1598 (2006).
97. Kishi, T. & Elmquist, J. K. Body weight is regulated by the brain: a link between feeding and emotion. *Mol. Psychiatry* **10,** 132–146 (2005).
98. Kaplitt, M. G. *et al.* Safety and tolerability of gene therapy with an adeno-associated virus (AAV) borne GAD gene for Parkinson's disease: an open label, phase I trial. *Lancet* **369,** 2097–2105 (2007).
This paper demonstrates the efficacy and safety of stereotactic viral-mediated gene therapy in the treatment of severe Parkinson's disease.
99. Cryan, J. F., Markou, A. & Lucki, I. Assessing antidepressant activity in rodents: recent developments and future needs. *Trends Pharmacol. Sci.* **23,** 238–245 (2002).

Acknowledgements Work in E.J.N.'s laboratory was supported by grants from the National Institute of Mental Health.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence should be addressed to E.J.N. (eric.nestler@mssm.edu).

Neuroligins and neurexins link synaptic function to cognitive disease

Thomas C. Südhof[1]

The brain processes information by transmitting signals at synapses, which connect neurons into vast networks of communicating cells. In these networks, synapses not only transmit signals but also transform and refine them. Neurexins and neuroligins are synaptic cell-adhesion molecules that connect presynaptic and postsynaptic neurons at synapses, mediate signalling across the synapse, and shape the properties of neural networks by specifying synaptic functions. In humans, alterations in genes encoding neurexins or neuroligins have recently been implicated in autism and other cognitive diseases, linking synaptic cell adhesion to cognition and its disorders.

The brain integrates and processes sensory inputs to generate motor outputs appropriate for the survival of the organism. This transformation of information is carried out by cascades of synapses, assembled in overlapping neural circuits[1]. All processing of information in the brain involves synapses, and almost all abnormalities in brain function have a direct or indirect effect on synaptic function.

Synapses are specialized intercellular junctions dedicated to the transfer of information from a neuron to a target cell, usually another neuron[1] (Fig. 1a). Synaptic transmission of information is fast, dynamic, efficient and tightly regulated (Box 1). Synapses have many of the same properties as intercellular junctions in non-neural tissues, but they differ from all other such junctions because they are inherently asymmetrical, transmit information by an extremely fast mechanism, and are highly plastic. Moreover, synapses have diverse properties that are specified by both the presynaptic neuron and the postsynaptic neuron (for example, see ref. 2).

Enormous progress has recently been made in understanding synaptic transmission; much is now known about the machinery and functional properties of synapses. However, the molecular mechanisms underlying synapse formation and the specification of synapse diversity are less clear, as are the processes mediating the assembly of synapses into neural circuits[3]. For appropriate neural-circuit function, the formation and specification of synapses is immensely important. The input–output properties of a neural circuit depend on both its pattern of synaptic connectivity (referred to as its wiring diagram) and the diverse properties of individual synapses in the circuit[4]. The pattern of connectivity in a circuit is no more important than the properties of the individual synapses comprising the circuit. Use-dependent changes in synaptic strength (that is, synaptic plasticity) can completely alter the relative contributions of different synapses in a circuit, thereby sometimes even reversing its input–output properties as a function of previous use without a change in the wiring diagram (for example, see ref. 5).

Synapse formation and the specification of synaptic diversity are intricately linked and probably depend on the actions of synaptic cell-adhesion molecules[3]. The diversity of synapses is partly due to differences in the composition of their neurotransmitter release and receptor machineries but seems to be based largely on differences in the organization of these machineries. Synapse formation and specification probably involves three steps: initial recognition of the target cell by the neural growth cone, formation of synaptic junctions with recruitment of synaptic components, and maturation of synaptic junctions with specification of circuit-specific properties. Functional assays for synapse formation and tests of specific molecules are difficult to carry out (Box 2), hindering identification of the molecular mechanisms involved. These difficulties are confounded by the fact that many candidate molecules, such as cadherins and WNTs, carry out essential functions during earlier development, in addition to their presumptive role in synapse formation[6,7].

Vertebrate neurexins (NRXNs) and neuroligins (NLGNs) are arguably the best-characterized synaptic cell-adhesion molecules, and they are the only ones for which a specifically synaptic function has been established[8,9]. Here I describe the role of NRXNs and NLGNs as synaptic cell-adhesion molecules that function in an unexpected manner. I suggest that these molecules are required for synapse function but not for synapse formation, that they affect trans-synaptic activation of synaptic transmission but are not essential for synaptic cohesion of the presynaptic and postsynaptic specializations, and that their dysfunction impairs the properties of synapses and disrupts neural networks without completely abolishing synaptic transmission[10–12]. Because they are cell-adhesion molecules, NRXNs and NLGNs probably function by binding to each other and by interacting with intracellular proteins (most notably with PDZ-domain proteins), but the precise mechanisms involved and their relationship to synaptic transmission remain unclear. The importance of NRXNs and NLGNs for synaptic function is, however, evident from the marked deficits in synaptic transmission in mice lacking NRXNs or NLGNs.

The role of NRXNs and NLGNs in synaptic function almost predestines them for a role in cognitive diseases, such as schizophrenia and autism spectrum disorders (ASDs), the mechanisms of which have proved difficult to ascertain. One reason for the difficulties in understanding cognitive diseases is that they may arise from subtle changes in a subset of synapses in a neural circuit, as opposed to a general impairment of all synapses in all circuits. As a result, the same molecular alteration can produce different circuit changes and neurological symptoms, which are then classified

[1]Neuroscience Institute, Department of Molecular and Cellular Physiology, and Howard Hughes Medical Institute, Stanford University, 1050 Arastradero Road B249, Palo Alto, California 94304, USA.

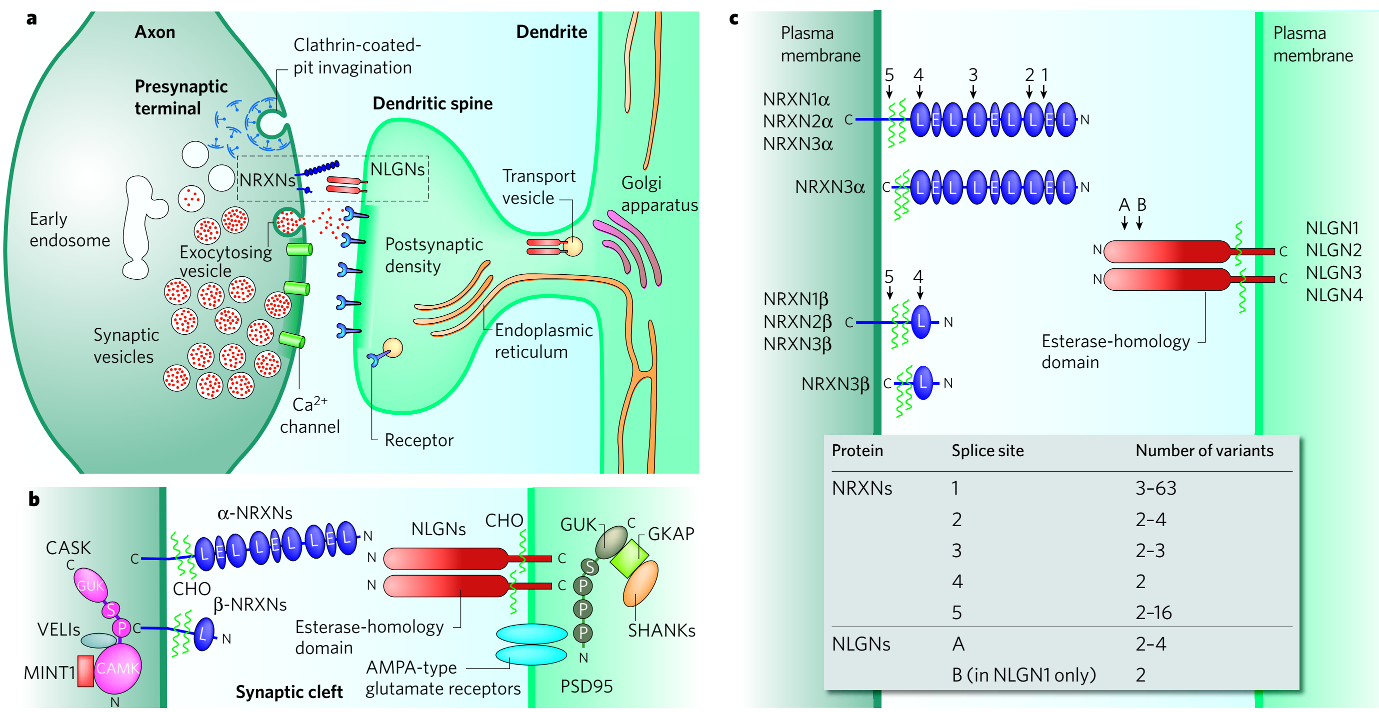
as distinct cognitive diseases. Indeed, recent studies have identified mutations in the genes encoding NRXNs and NLGNs as a cause of ASDs, Tourette's syndrome, learning disability and/or schizophrenia; sometimes family members with the same mutation have different cognitive disorders[13–27]. Viewed as a whole, current results thus identify NRXNs and NLGNs as trans-synaptic cell-adhesion molecules that mediate essential signalling between presynaptic and postsynaptic specializations. This signalling is central to the brain's ability to process information and is a key target in the pathogenesis of cognitive diseases. I discuss the idea that identification of mutations in NRXNs and NLGNs in patients with cognitive diseases, especially ASDs and schizophrenia, supports the notion that these diseases are caused, at least in part, by abnormalities in synaptic transmission in a subset of neural circuits.

Neurexins are polymorphic synaptic receptors

The venom of the black-widow spider contains a vertebrate-specific toxin called α-latrotoxin. α-Latrotoxin is a large protein that binds to presynaptic receptors and induces a massive release of neurotransmitters[28]. NRXNs were discovered as receptors for α-latrotoxin[29]. NRXNs are type I membrane proteins and can be classified into two types: α-NRXNs and β-NRXNs. α-NRXNs are larger than β-NRXNs; the two types of NRXN contain different amino-terminal extracellular sequences but identical carboxy-terminal transmembrane regions and cytoplasmic tails (Fig. 1b). Extracellularly, α-NRXNs have six LNS domains (laminin, NRXN, sex-hormone-binding globulin domains) with three intercalated epidermal growth factor (EGF)-like domains, whereas β-NRXNs have a single LNS domain. In addition to α-NRXNs and β-NRXNs, neurons express NRXN-related proteins called CASPRs (contactin-associated proteins), which resemble α-NRXNs but contain additional extracellular domains that are not found in α-NRXNs[30]. CASPRs also function as cell-adhesion molecules, like NRXNs, but are mainly involved in neuron–glia interactions outside synapses[31].

The mammalian genome contains three NRXN genes (*NRXN1*, *NRXN2* and *NRXN3*), each encoding an α-protein and a β-protein from independent promoters[32]. Furthermore, extensive alternative splicing of the encoded proteins at five canonical positions generates thousands of NRXN isoforms[33] (Fig. 1c). Conceptually, these isoforms could specify a 'code' of interactions at synapses. Consistent with this idea, alternative splicing of NRXNs is regionally regulated and altered by activity in neurons[33,34]. Splice sites 1 to 4 (SS 1 to SS 4) involve short sequences (30 residues or fewer), are located in or adjacent to LNS domains and are conserved in all three NRXNs. Splice site 5 in *NRXN1* involves the insertion of only three residues. But in *NRXN2*, splicing at this site inserts 191 residues, and in *NRXN3* it creates a huge diversity of sequence inserts that include multiple variants with in-frame stop codons and therefore encode secreted NRXNs[33,35].

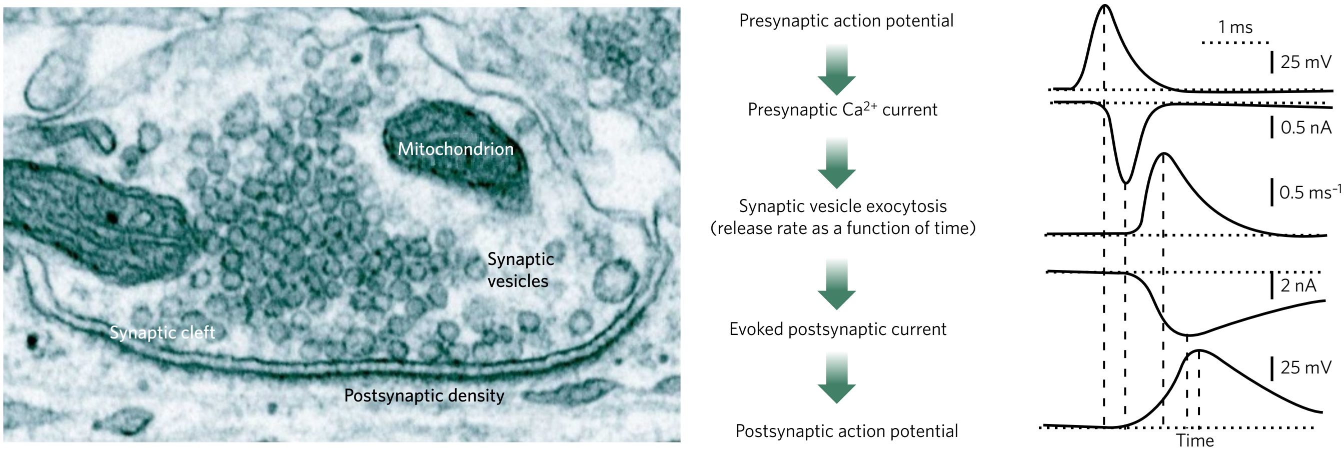
Using *in situ* hybridizations, messenger RNAs encoding the different α-NRXNs and β-NRXNs were shown to be coexpressed in the same class of neuron, but each type of NRXN was found to be differentially distributed between different classes of neuron[33]. Immunofluorescence studies, subcellular fractionations and the function of NRXNs as α-latrotoxin receptors indicate that NRXNs are located on presynaptic terminals[29,36–38]. It remains unclear, however, whether NRXNs are



Protein	Splice site	Number of variants
NRXNs	1	3-63
	2	2-4
	3	2-3
	4	2
	5	2-16
NLGNs	A	2-4
	B (in NLGN1 only)	2

Figure 1 | Architecture of the trans-synaptic neurexin–neuroligin complex. a, The structure of an excitatory synapse and the putative locations of neurexins (NRXNs) and neuroligins (NLGNs) in the synapse. A presynaptic varicosity containing synaptic vesicles is shown on the left, and a postsynaptic spine with a postsynaptic density containing neurotransmitter receptors on the right. **b**, The NRXN–NLGN junction, including selected presynaptic and postsynaptic binding proteins: CASK, VELIs and MINTs on the presynaptic side[62]; and PSD95 (which binds to AMPA-type glutamate receptors through its first PDZ domain, and to NLGNs through its third PDZ domain[64]), GKAP and SHANKs on the postsynaptic side. Note that a proportion of NRXNs and CASK could be also postsynaptic, and that SHANKs might also be presynaptic. C, carboxyl terminus; CAMK, Ca^{2+}/calmodulin-dependent protein kinase domain of CASK; CHO, carbohydrate-attachment sequence; E, epidermal growth factor (EGF)-like domain; GUK, guanylate-kinase domain; L, LNS (laminin, NRXN, sex-hormone-binding globulin) domain; N, amino terminus; P, PDZ domain; S, SH3 domain. **c**, Alternative splicing of NRXNs and NLGNs. α-NRXNs contain five canonical splice sites (1 to 5), and β-NRXNs contain two (4 and 5). Splice site 1 is C-terminal to the first EGF-like domain. Splice sites 2, 3 and 4 are at similar positions in the second, fourth and sixth LNS domains, respectively. Splice site 5 is between the glycosylated CHO sequence and the transmembrane region. Most alternative splicing involves the insertions of small evolutionarily conserved sequences, except for at splice site 5. Splicing at this site in NRXN2 involves a large insertion (191 residues), and in NRXN3 it involves at least 16 variants, some of which include stop codons and thus produce secreted isoforms of NRXN3 (ref. 35). NLGNs contain only two sites of alternative splicing, and one of these sites, site B, is present only in NLGN1.

Box 1 | How a synapse works



At a synapse, a presynaptic terminal containing abundant synaptic vesicles contacts a postsynaptic cell, usually a neuron, as shown in the electron micrograph of a synapse in the left panel above. When an action potential invades the presynaptic terminal, Ca^{2+} channels open, and the inflowing Ca^{2+} triggers fusion of synaptic vesicles with the presynaptic plasma membrane, thereby emptying the neurotransmitters in the vesicles into the synaptic cleft[90]. The neurotransmitters then bind to postsynaptic receptors to complete the information transfer. The overall process is incredibly rapid, with each of the main steps (presynaptic vesicle fusion and postsynaptic signal reception) initiating in less than 1 ms. The time course of synaptic transmission as measured electrophysiologically is shown in the right panel above. The five sequential steps are indicated, as deduced from measurements in the rat calyx of Held synapse[90]. Vertical dashed lines identify the time points when the five steps described reach their peak; horizontal dashed lines identify baselines. In addition to this classic mode of synaptic transmission, synapses have other types of signalling that operate on a longer timescale and function to regulate the synaptic transmission. Structurally, synapses are characterized by coats that line the intracellular face of the presynaptic plasma membrane (referred to as the active zone because synaptic vesicles undergo fusion here) and the postsynaptic plasma membrane (referred to as the postsynaptic density). Presynaptic and postsynaptic plasma membranes are always precisely aligned, and they are separated by a synaptic cleft of about 20 nm. The cleft contains an undefined proteinaceous material in the middle and is presumably bridged by synaptic cell-adhesion molecules such as neurexins (NRXNs) and neuroligins (NLGNs) that align the presynaptic and postsynaptic elements and mediate trans-synaptic signalling. (Left panel courtesy of X. Liu, University of Texas Southwestern Medical Centre, Dallas. Right panel modified, with permission, from ref. 91.)

confined to the presynaptic side, because deletion of genes encoding α-NRXNs also has postsynaptic effects[39] and a proportion of NRXNs has been shown to be present on postsynaptic sites[40].

Neuroligins are neurexin ligands

NLGNs are endogenous NRXN ligands[41]. They are type I membrane proteins, like NRXNs, but have a simpler domain structure and are less diverse (Fig. 1b, c). In addition to NLGNs, neurexophilins (neuropeptide-like proteins) and dystroglycan (a cell-adhesion molecule involved in many types of junction) are also NRXN ligands[42,43]. In contrast to NLGNs, however, no functional effect of binding of neurexophilin or dystroglycan to NRXNs has been observed so far.

The extracellular sequences of NLGNs are composed of a single domain that is homologous with acetylcholinesterases, but it lacks crucial residues in the active site, which is thus disabled (Fig. 1b, c). NLGNs form constitutive dimers through this domain, which is connected to the single transmembrane region by a glycosylated linker sequence. Mammals express four genes encoding NLGNs, with *NLGN3* and *NLGN4* in humans localized to the X chromosome. In humans, *NLGN4* is complemented on the Y chromosome by a similar gene, *NLGN5*. All NLGNs are alternatively spliced at a single canonical position (referred to as SS A); in addition, NLGN1 is alternatively spliced at a second position (called SS B)[44,45]. Most NRXNs and NLGNs are evolutionarily conserved in vertebrates and have more distant relatives in invertebrates[32,46]. It is of interest that the gene encoding NLGN4 diverged rapidly in rodents, suggesting that at least some genes encoding NLGNs are subject to less evolutionary constraint[47]. Sequence comparisons indicate that NLGN1, NLGN3 and NLGN4/NLGN5 are more similar to each other than to NLGN2. All NLGNs are enriched at postsynaptic densities, as judged by subcellular localization. Immunocytochemistry revealed that NLGN1 and NLGN2 are exclusively localized to excitatory and inhibitory synapses, respectively, whereas NLGN3 might be present in both[48–51].

NLGNs bind to both α-NRXNs and β-NRXNs with nanomolar affinities; binding involves the sixth LNS domain of α-NRXNs, which corresponds to the only LNS domain of β-NRXNs[52]. The binding affinities differ characteristically between various pairs of NLGNs and NRXNs, and they are controlled by alternative splicing of both NRXNs and NLGNs[45,52,53] (Fig. 1c). SS B of NLGN1 is a master switch for NRXN binding — the inclusion of only eight residues in this site restricts the binding of NLGN1 to those β-NRXNs that lack an insert in SS 4, whereas exclusion of these eight residues allows the binding of both α-NRXNs and β-NRXNs independently of SS 4 (ref. 45). The NLGN1 splice variant containing an insert in SS B predominates, indicating that most NLGN1 is specific for β-NRXNs lacking an insert in SS 4, whereas all other NLGNs bind to both α-NRXNs and β-NRXNs. SS A of all NLGNs also regulates NRXN binding, but the effect is smaller[52]. In NRXNs, SS 4 (which is located in the last LNS domain) not only controls the binding of β-NRXNs to NLGN1 containing an insert in SS B (discussed earlier) but also modulates the affinity of α-NRXNs and β-NRXNs for NLGNs lacking an insert in SS B. Thus, the current data suggest that NRXN–NLGN binding is governed by a hierarchical code that depends on which principal isoforms are expressed and which splice variants are used.

The trans-synaptic neurexin–neuroligin complex

NRXNs and NLGNs are thought to form a trans-synaptic complex that is coated on both sides by PDZ-domain-containing proteins (Fig. 1b). The crystal structure of the NRXN1–NLGN1 complex (without inserts in SS 4 of NRXN and SS B of NLGN1) revealed that the NRXN LNS domain attaches with a large contact area to the lateral sides of the NLGN esterase-homology domain, opposite to the position of the crippled active site[54–56] (Fig. 2). In the structure of crystals that were grown in the presence of Ca^{2+}, two fully occupied Ca^{2+}-binding sites were found that are coordinated by ligands from both proteins[55]. Mapping of the alternative splicing sites into the structure shows that SS B

Box 2 | Analysing synaptic cell-adhesion molecules and synapse formation

Studying synaptic cell-adhesion molecules functionally has turned out to be extremely difficult. This box lists the assays used to study these molecules and summarizes the advantages and the disadvantages associated with each assay.

Gain-of-function approaches

Cell-adhesion assay
This type of assay uses non-neuronal cells expressing cell-adhesion molecules to test whether these molecules can mediate stable cell–cell interactions (for example, between NRXNs and NLGNs) but provides no functional read-out.

Artificial synapse formation assay
In this assay, neurons are cultured together with non-neuronal cells expressing a cell-adhesion molecule. The assay tests whether the cell-adhesion molecule induces the neurons to form stable junctions with synapse-like properties with the non-neuronal cells[38,50,67,68]. Many molecules promote synapse formation in the artificial synapse formation assay.

Neuronal transfection assay
This assay uses neurons overexpressing a cell-adhesion molecule. The assay measures the synapse density on the transfected neurons by microscopy[53] and the synapse function by electrophysiology[12]. The assay allows a better functional analysis of the effects of a cell-adhesion molecule than the artificial synapse formation assay; however, neither directly measures synapse formation, and both are subject to overexpression artefacts.

Loss-of-function approaches

RNA interference
This type of experiment uses cultured neurons or cultured slices and tests whether a cell-adhesion molecule is essential for synapse formation or synapse function. When paired with rescue controls, RNA interference (RNAi) is ideal, but it has three potential limitations. First, it is difficult to target multiple proteins simultaneously with RNAi, and it is therefore difficult to address redundancy. Second, for many targets, RNAi is simply inefficient; that is, it achieves less than 75% suppression when measured quantitatively (and not by densitometry of blots). Even successful RNAi is never complete — it does not achieve more than 95% suppression. Third, compensatory changes are as likely during RNAi-mediated knockdown experiments as during gene-knockout experiments.

Constitutive genetic manipulation
Constitutive genetic manipulation in gene-targeting experiments permanently deletes or alters the expression of a gene to test its overall importance. In addition to the problems listed for RNAi, this approach has the potential to cause developmental alterations, but it allows the complete elimination of expression and makes organismal analyses possible.

Conditional deletion
Conditional deletion by means of gene targeting allows spatially and/or temporally regulated deletion or changes in a cell-adhesion molecule, and usually involves Cre-recombinase-mediated genetic changes. This is a powerful approach, but it is labour intensive and is limited by the paucity of mouse lines with reproducible, tight and robust expression of Cre recombinase.

Pharmacological inhibition
This type of inhibition of a cell-adhesion molecule can be used to cause an acute disruption of function. Potentially the best approach, it is limited by lack of availability of effective agents for almost all cell-adhesion molecules, and by the side effects of many of the agents that do exist.

Overall evaluation
Gain-of-function approaches for analysing synapse formation are more sensitive but are harder to interpret. Loss-of-function approaches have greater validity but are technically more difficult, and they can be limited by functional redundancy between multiple genes. Note that both gain-of-function approaches and loss-of-function approaches, including RNAi and overexpression experiments, suffer from the potential problem of compensatory changes in the expression, localization and/or stability of other proteins.

is included in the binding interface, and that SS A of NLGN1 and SS 4 of NRXN1 are close by, providing an explanation for the effect of alternative splicing of these sites on the NRXN–NLGN binding affinity. Indeed, direct comparison of the crystal structures of β-NRXN LNS domains containing and lacking inserts in SS 4 supports this conclusion by revealing major conformational changes induced by this alternative splicing event[57,58].

The shape of the NRXN–NLGN complex suggests that it forms an interaction layer in the centre of the synaptic cleft, with the C-terminal sequences emerging from the complex in opposite directions (Fig. 2). This interaction layer — which might contribute to the electron-dense material observed in the synaptic cleft by electron microscopy — is separated from the presynaptic and postsynaptic plasma membranes by the glycosylated linker sequences that are present in NRXNs and NLGNs just outside the membrane. These glycosylated sequences could function as a 'cuff' that creates a distance between the interaction layer and the plasma membranes, and forces the extracellular domains to project into the synaptic cleft away from the membrane.

The cytoplasmic sequence of NRXNs contains a C-terminal binding site for class II PDZ domains that binds to the PDZ domain of CASK and related proteins, and a membrane-proximal binding site for protein 4.1 (refs 59, 60). CASK is a MAGUK protein (for membrane-associated guanylate kinase protein) containing a PDZ domain, an SH3 domain and a guanylate-kinase domain. CASK is an unusual MAGUK, however, because the PDZ, SH3 and guanylate-kinase domains account for only its C-terminal half; its N-terminal half is occupied by a Ca^{2+}/calmodulin-dependent protein kinase (CAMK) domain that is absent from other MAGUKs. The CASK CAMK domain contains substitutions in canonical residues that coordinate Mg^{2+} in CAM kinases,suggesting that it may be catalytically inactive. Unexpectedly, however, recent evidence indicates that CASK is catalytically active, and may be the first described kinase that phosphorylates proteins, including NRXN1, independently of Mg^{2+} (ref. 61). In addition, CASK nucleates the assembly of actin on the cytoplasmic sequence of NRXN by simultaneously binding to protein 4.1 (ref. 60). Finally, CASK interacts with VELI proteins (also known as MAL proteins, which are homologues of *Caenorhabditis elegans* LIN-7) and with MINT proteins (also known as X11 proteins) to form a tight trimeric complex[62,63]. In addition to NRXNs, CASK binds to other cell-surface proteins, including CASPRs, and probably carries out analogous functions. Deletion of *Cask* in mice causes a lethal phenotype that includes synaptic abnormalities, indicating that CASK is an important molecule[64]. CASK is probably a component of a signal transduction cascade that translates extracellular interactions of cell-surface proteins into an intracellular response by modulating the actin cytoskeleton and phosphorylating target proteins.

Like NRXNs, NLGNs bind to intracellular PDZ-domain proteins, but in contrast to NRXNs, NLGNs bind to class I PDZ domains such as those contained in PSD95, a postsynaptic MAGUK protein[65]. PSD95 and its homologues are centrally involved in recruiting glutamate receptors at postsynaptic sites[66]. Similarly to CASK, PSD95 binds to intracellular adaptor proteins, and especially to GKAP (a protein that binds to the guanylate-kinase domain of PSD95), which, in turn, binds to SHANK proteins (Fig. 1b). A possible role of these interactions is to recruit post-synaptic adaptor proteins to the site of synaptic junctions. As a result of their binding to PDZ-domain proteins, the junction formed by NRXNs and NLGNs resembles the architecture of tight junctions, but it differs from them in that the NRXN–NLGN junction is asymmetrical in all of its components.

Function of neuroligins and neurexins

Initial evidence that NLGNs function at synapses came from ingenious experiments showing that NLGNs expressed in a non-neuronal cell can induce co-cultured neurons to form presynaptic specializations onto the non-neuronal cell[67] (Box 2). This finding was amplified by complementary experiments showing that NRXNs, when expressed in a non-neuronal cell, can induce the formation of postsynaptic specializations in co-cultured neurons[50,68]. Moreover, direct overexpression of NLGNs in transfected neurons caused an increase in synapse numbers on these neurons[69].

Taken together, these studies indicated that NLGNs and NRXNs may induce synapse formation. However, analysis of gene-knockout mice surprisingly revealed that NLGNs and α-NRXNs are essential for synaptic function, but not for synapse formation[10–12]. Mice lacking NLGN1, NLGN2 and NLGN3 die at birth, but they have nearly normal synapse numbers with an apparently normal ultrastructure. Electrophysiological analyses in sections cut from brains of the mutant mice and analysed immediately without culturing showed that these mice have a severe impairment of synaptic transmission[11]. Although mice that lack the gene encoding either NLGN1 or NLGN2 are viable and fertile, electrophysiological analysis also uncovered significant synaptic dysfunctions in these mice[12]. In agreement with the localizations of NLGN1 and NLGN2 to excitatory and inhibitory synapses, respectively, excitatory synapses showed impairments in *N*-methyl-D-aspartate (NMDA)-receptor signalling in *Nlgn1*-knockout mice, whereas *Nlgn2*-knockout mice had deficits in inhibitory synaptic transmission[12].

The gene-knockout analysis seems to contradict the *in vitro* assays showing that NLGNs induce synapses in the artificial synapse formation assay and the neuronal transfection assay (see Box 2, which explains the various approaches). However, the assays with cultured neurons do not directly measure synapse induction — rather, they measure an increase in synapse numbers after a particular manipulation. In these assays, the molecules tested could simply function by inducing signal transduction events that stabilize otherwise transient, tentative synaptic contacts. In support of this interpretation, and in agreement with the results from gene-knockout experiments, the ability of NLGNs to increase the number of synapses in a transfected neuron can be decreased by the inhibition of synaptic activity, which has no effect on the expression and localization of the transfected NLGNs[12]. More conclusively, paired recordings from inhibitory neurons in the somatosensory cortex of either *Nlgn1*- or *Nlgn2*-knockout mice demonstrated that deletion of the *Nlgn1* and *Nlgn2* did not decrease the number of synaptic connections (Fig. 3). Instead, deletion of *Nlgn2* (but not *Nlgn1*) selectively lowered the strength of GABA (γ-aminobutyric acid)-utilizing inhibitory synapses formed by fast-spiking, parvalbumin-containing interneurons, but not of GABA-utilizing synapses formed by somatostatin-containing interneurons. Together, these data suggest that NLGNs have a function in the maturation of synaptic junctions with the specification of circuit-specific properties, but not in the initial formation of synaptic junctions. This conclusion is consistent with the finding that a partial knockdown of the mRNAs encoding NLGNs in cultured neurons produced a partial decrease in synapse numbers that could have been a secondary consequence of a decrease in synaptic function[70].

The activities of NRXNs have been more difficult to characterize than those of NLGNs. The lack of high-affinity antibodies, the complexity of the NRXN isoforms, and the challenges in analysing presynaptic function have contributed to this difficulty. At present, it even remains uncertain whether NRXNs are exclusively presynaptic, or whether at least some NRXNs are postsynaptic. Analysis of mice that lacked all α-NRXNs but still had β-NRXNs uncovered a phenotype that is similar to that of the *Nlgn*-knockout mice described above (note that knockout mice lacking β-NRXNs or both α-NRXNs and β-NRXNs have not yet been analysed). Deletion of the genes encoding individual α-NRXNs causes only moderate increases in mortality in mice, but deletion of two of the three genes encoding α-NRXNs increases postnatal mortality markedly. Deletion of all three genes invariably leads to neonatal fatality in mice[10]. Again, synapse numbers and their ultrastructure

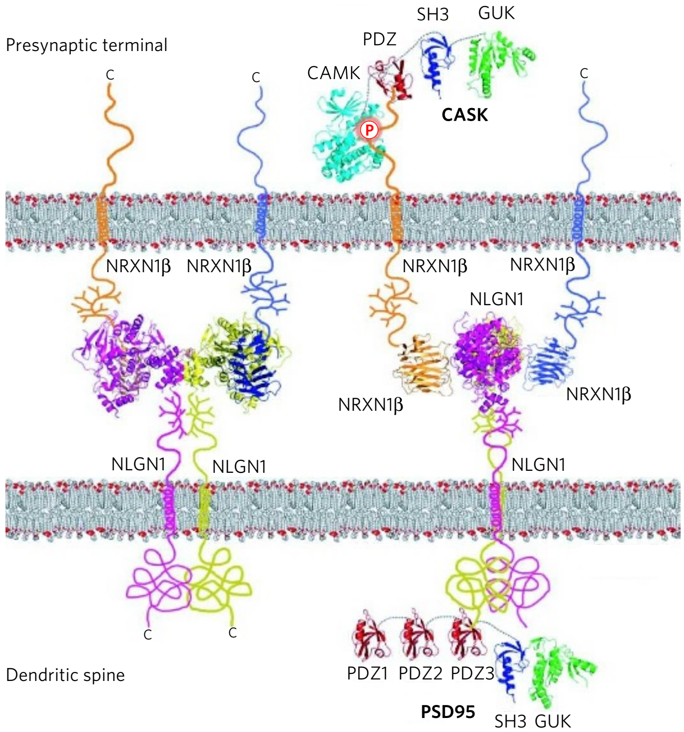
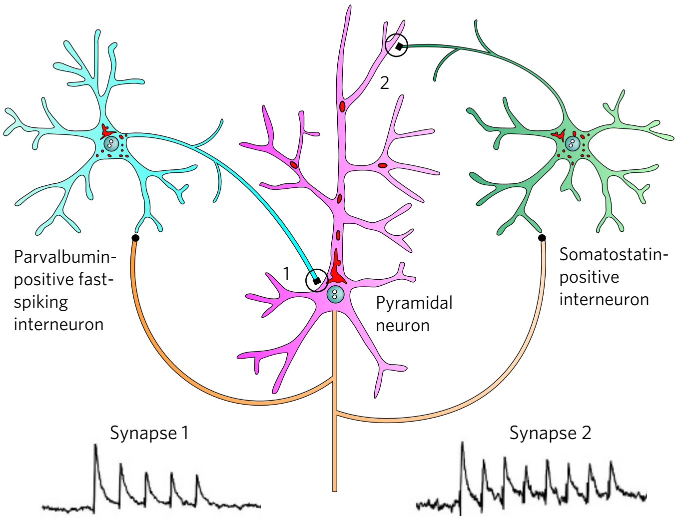


Figure 2 | Atomic model of the trans-synaptic complex formed by NRXN1β and NLGN1. The NRXN1β–NLGN1 complex is shown in two orientations: left, en face with the NRXN1β LNS domain on top of the NLGN1 esterase-like domain, to illustrate the NLGN1 dimer; right, in a 90° rotation to illustrate the sideways attachment of the NRXN1β LNS domains to the NLGN1 esterase-like domains. The two orientations of the NRXN1β–NLGN1 complex illustrate the spatial arrangement and relative sizes of the NRXN LNS domains and the NLGN esterase-like domain in a synaptic cleft. Other NRXN and NLGN isoforms for which no full structure is available, including α-NRXNs, would presumably have a similar arrangement except that the additional LNS domain in α-NRXNs would occupy a larger space in the synaptic cleft. (Images were generated from the following Protein Data Bank files: NRXN–NLGN, 3BIW; PSD95 SH3 GUK, 1JXO; PSD95 PDZ, 1BE9; CASK CAMK, 3C0I; CASK GUK, 1KGD; CASK PDZ, 1KWA. Diagram courtesy of D. Arac and A. Brunger, Stanford University, California.)

are nearly normal in mice that lack α-NRXNs, but synapse function is severely impaired. This impairment is both presynaptic and postsynaptic but is most significantly observed in action-potential-driven neurotransmitter release, which is severely depressed, largely because of a loss of presynaptic Ca^{2+}-channel function[71]. Postsynaptically, deletion of α-NRXN-encoding genes caused a decrease in synaptic responses dependent on NMDA receptors but not on AMPA receptors, similarly to the deletion of *Nlgn1* (ref. 39). Overall, analysis of mice in which α-NRXN-encoding genes have been knocked out indicates that lack of these genes results in disorganized synapses. These findings characterize α-NRXNs, like NLGNs, as synaptic cell-adhesion molecules that are essential for the proper assembly of synapses into a fully functional unit but not for the initial formation of synapses. NRXNs may also be required globally for the organization of secretory systems because mice that lack α-NRXNs show an additional major impairment in neuroendocrine secretion[72].

How, precisely, do NRXNs and NLGNs function in synapses? A plausible hypothesis is that trans-synaptic cell adhesion mediated by NRXNs and NLGNs — either by binding to each other or by binding to other ligands — triggers presynaptic and postsynaptic signal transduction events that activate synaptic function and specify synaptic properties. Without this activation, synapses assemble but do not work properly. The activation is clearly not a simple yes-or-no switch. Instead, NRXNs and NLGNs shape synaptic efficacy and plasticity. Moreover, a synaptic



Knockout mouse	Synapse 1		Synapse 2	
	Connectivity	IPSC amplitude	Connectivity	IPSC amplitude
Nlgn1	↔	↔	↔	↔
Nlgn2	↔	↓	↔	↔
Nlgn1 and *Nlgn2*	↔	↓	↔	↔

Figure 3 | Differential effects of deletion of the gene encoding NLGN1 and NLGN2 on inhibitory synapses in the somatosensory cortex. Connections of parvalbumin-positive fast-spiking interneurons (blue) and of somatostatin-containing interneurons (green) with excitatory pyramidal neurons (pink) are shown schematically. The connectivity (measured in paired recordings as percentage success) and amplitude of the inhibitory synapses of the interneurons onto the pyramidal neuron (IPSC amplitude) was measured for *Nlgn1*-knockout mice, *Nlgn2*-knockout mice, and *Nlgn1* and *Nlgn2* double-knockout mice; the table shows how these values differ from the connectivity and amplitude in wild-type mice (J. Gibson, K. Huber and T.C.S., unpublished observations). Horizontal arrow, no change compared with wild-type mice; downward arrow, lower than in wild-type mice.

transmission-specific element is involved, at least for NLGNs. How this synapse activation may occur is unclear. Binding of NLGN to NRXNs does not induce the dimerization of NRXNs, in contrast to other receptor-dimerization-dependent signalling cascades, because the crystal structure reveals that the two NRXNs bound to a NLGN dimer are distant monomers[54–56] (Fig. 2). The most parsimonious mechanism for this activation would be that NRXNs and NLGNs recruit 'coats' to the junction, coats that might consist of PDZ-domain proteins or actin filaments, or might also involve other types of interaction. For example, binding of NLGN to NRXNs might stimulate the CASK-dependent phosphorylation of NRXNs and other substrates; however, no direct evidence for this mechanism exists.

Neuroligins and neurexins in autism

ASDs are common and enigmatic diseases. They comprise classical idiopathic autism, Asperger's syndrome, Rett's syndrome, and pervasive developmental disorder not otherwise specified[73,74]. Moreover, several other genetic disorders, such as Down's syndrome, fragile X mental retardation, and tuberous sclerosis, are frequently associated with autism. Such syndromic forms of autism and Rett's syndrome are usually more severe because of the nature of the underlying diseases. The key features of ASDs are difficulties in social interactions and communication, language impairments, a restricted pattern of interests, and/or stereotypic and repetitive behaviours. Learning disability (in about 70% of cases) and epilepsy (in about 30% of cases) are frequently observed; in fact, the observation of epilepsy in patients with ASDs has fuelled speculation that autism may be caused by an imbalance of excitatory and inhibitory synaptic transmissions. In rare instances, idiopathic autism is associated with specialized abilities, for example in music, mathematics or memory. The relationship of ASDs to other cognitive diseases such as schizophrenia and Tourette's syndrome is unclear. As is the case for phenotypes caused by mutations in NLGNs and NRXNs (discussed later), the boundaries between the various disorders may not be as real as the clinical manifestations suggest.

A key feature of ASDs is that they typically develop before two or three years of age[73,74]. They therefore affect brain development relatively late, during the time of human synapse formation and maturation. Consistent with this time course is the fact that few anatomical changes are associated with ASDs[75]. An increase in brain size has repeatedly been reported[76] but is not generally agreed on[75]. Thus, similarly to other cognitive diseases, ASDs are a disorder not of brain structure but of brain function. Among cognitive diseases, ASDs are the most heritable (about 80%), suggesting that they are determined largely by genes and not by the environment. ASDs show a male-to-female ratio of about 4:1, indicating either that ASDs involve the X chromosome directly or that the penetrance of pathogenic genes is facilitated in males[73,74].

Mutations in many genes have been associated with familial ASDs. A consistent observation emerging from recent studies is the discovery of mutations in the genes encoding NRXN1, NLGN3 and NLGN4. Specifically, seven point mutations, two distinct translocation events and four different large-scale deletions in the *NRXN1* gene were detected in patients with autism[13–18]. Ten different mutations in the *NLGN4* gene were observed (two frameshifts, five missense mutations and three internal deletions), and a single mutation in the *NLGN3* gene (the Arg451Cys substitution)[21–24]. Besides these mutations, five different larger deletions of X-chromosomal DNA that includes the *NLGN4* locus (referred to as copy-number variations) were detected in patients with autism[18,25–27].

In addition to the NRXN–NLGN complex, mutations in the gene encoding SHANK3 — an intracellular scaffolding protein that binds indirectly to NLGNs through PSD95 and GKAP (ref. 66; Fig. 1) — may also occur frequently in ASDs. An astounding 18 point mutations were detected in the *SHANK3* gene in patients with autism, in addition to several cases containing copy-number variations that cover the gene[18,77–82]. Indeed, terminal 22q deletion syndrome is a frequent occurrence that shows autistic features that have been correlated with the absence of *SHANK3*, which is normally localized to this chromosome. SHANK3 is particularly interesting because it not only interacts indirectly with NLGNs but also binds directly to CIRLs (latrophilins), which, in turn, constitute α-latrotoxin receptors similar to NRXNs, suggesting a potential functional connection between SHANK3 and NRXNs[83].

Overall, the description of the various mutations in the NRXN–NLGN–SHANK3 complex seems to provide overwhelming evidence for a role of this complex in ASDs, given the fact that, in total, these mutations account for a significant proportion of patients with autism. It should be noted, however, that two issues give rise to scepticism with regard to the role of this complex in ASDs.

First, at least for some of the mutations in this complex, non-symptomatic carriers were detected in the same families in which the patients with the mutations were found. Whereas the *NLGN3* and *NLGN4* mutations seem to be almost always penetrant in males, and even female carriers with these mutations often have a phenotype, the *SHANK3* point mutations in particular were often observed in non-symptomatic siblings[77,78]. Thus, these mutations may only increase the chance of autism rather than actually causing it.

Second, the same mutations can be associated with different phenotypes in different people. For example, a microdeletion in *NLGN4* was found to cause severe autism in one brother but Tourette's syndrome in the other[26]. This raises the issue of whether the 'autism' observed in patients with mutations in these genes is actually autism, an issue that could also be rephrased as the question of whether autism is qualitatively distinct from other cognitive diseases, as opposed to being a continuum of cognitive disorders. In support of the latter idea, two different deletions in NRXN1 that are in the region encoding NRXN1α but not NRXN1β have also been observed in families with schizophrenia[19,20],

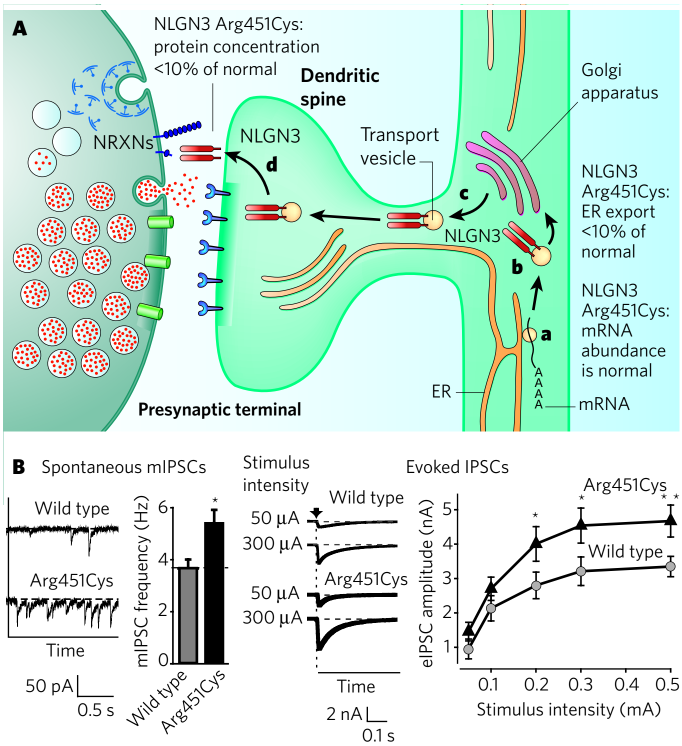


Figure 4 | The Arg451Cys substitution in NLGN3 impairs NLGN3 synthesis but enhances inhibitory synaptic transmission. A, The effect of the Arg451Cys mutation on NLGN3 synthesis. The mutation does not alter the amount of *Nlgn3* mRNA (**a**) but decreases the export of NLGN3 from the endoplasmic reticulum (ER) (**b**)[85]. As a result, the concentration of Arg451Cys mutant NLGN3 that is exported from the Golgi apparatus (**c**) and inserted into synapses (**d**) is less than 10% of the wild-type NLGN3 concentration[86]. **B**, Despite resulting in a decrease in the concentration of NLGN3, the Arg451Cys mutation produces a synaptic gain-of-function effect in inhibitory synapses in the somatosensory cortex. The figure illustrates, with two examples, an increased inhibitory synaptic activity in Arg451Cys mutant mice: by measurements of spontaneous 'miniature' synaptic events (mIPSCs; left) and by measurements of evoked synaptic responses (eIPSCs; right). Each example depicts representative electrophysiological traces and a summary graph. Asterisks indicate statistical significance (*P <0.05; **P <0.01). Note that NLGN3-deficient synapses do not show this phenotype[86]. IPSC, inhibitory postsynaptic current. (Panel modified, with permission, from ref. 86.)

indicating that there is a continuum of disorders that involve dysfunctions in synaptic cell adhesion and are manifested in different ways. Conversely, markedly different molecular changes can produce similar syndromes, as exemplified by the different mutations that are associated with ASDs[84].

At present, the relationship between the NRXN–NLGN synaptic-cell-adhesion complex and ASDs is tenuous. On the one hand, many of the mutations observed in familial ASD are clearly not polymorphisms but are deleterious, as demonstrated by the effect of these mutations on the structure or expression of the corresponding genes, and by the severe autism-like phenotypes observed in *Nlgn3* and *Nlgn4* mutant mice[85–87]. On the other hand, the nonlinear genotype–phenotype relationship in humans, evident from the only 70–80% heritability and from the occasional presence of mutations in non-symptomatic individuals, requires explanation. Work to explain the underlying mechanisms for this incomplete genotype–phenotype relationship is a promising avenue to insight into the genesis of autism. Furthermore, in addition to the link between schizophrenia and mutations in the gene encoding NRXN1α[19,20], linkage studies have connected NRXN3 to different types of addiction[88,89]. It is possible that, because of the nature of their function, mutations in genes encoding NRXNs and NLGNs constitute hotspots for human cognitive diseases.

Dissecting autism in mouse models

One way to address the question of whether the mutations in NRXNs and NLGNs observed in human patients are directly associated with ASDs is to test whether the same mutations elicit a significant phenotype in an animal. Such experiments were carried out in mice for two mutations in NLGNs: the NLGN3 Arg451Cys substitution and the NLGN4 loss-of-function mutation[86,87].

The Arg451Cys knock-in mouse has a striking phenotype with some, but not all, of the same features (as far as is possible to analyse) as human patients with ASDs. Behaviourally, the mice show normal motor and anxiety behaviours, a moderate impairment in social interactions, and a large increase in spatial learning capability[86]. Although this behavioural phenotype is somewhat satisfying because it is reminiscent of the 'savant' variant of autism and indicates that the Arg451Cys substitution did not impair cognitive function in the mice, the phenotype is also puzzling because human patients with the Arg451Cys substitution suffer from learning disabilities[21]. Electrophysiologically, the Arg451Cys mutant mice showed an increase in inhibitory synaptic transmission in the somatosensory cortex, which is consistent with the idea that a change in the excitatory–inhibitory balance contributes to the phenotype (Fig. 4). The Arg451Cys mutation seems to be a gain-of-function and not a loss-of-function mutation because *Nlgn3*-knockout mice did not have any of the phenotypes associated with the Arg451Cys knock-in mice[86]. This is surprising, because the Arg451Cys mutation depressed NLGN3 protein abundance in the knock-in mice by about 90%; thus, it is the remaining 10% of the mutant protein that produced a marked change in synaptic transmission (Fig. 4).

The gain-of-function action of the Arg451Cys mutation differs from that of the *Nlgn4* deletion, which also caused an autism-like phenotype in knockout mice[87], but it is clearly a loss-of-function mutation. These observations may provide an explanation for the finding of multiple mutations in *NLGN4* in patients with autism, but only a single mutation in *NLGN3*, despite the fact that both genes are on the X chromosome. It seems likely that only a loss-of-function mutation of *NLGN4*, but not of *NLGN3*, produces autistic symptoms, and that the Arg451Cys mutation in NLGN3 was an accidental gain-of-function mutation that occurred in only a single family. Moreover, these observations provide further support for the idea that NLGNs and NRXNs are activators of synapse function, not simply building blocks of synapses, in which small changes in NLGN function can induce massive changes in the neural network.

Perspective

Discovery of the NRXN–NLGN cell-adhesion system opened up new avenues to the understanding of synapses and cognitive disease, but it also raised many new questions. For example, do NRXNs and NLGNs function only by binding to each other — in fact, do they actually function by binding to each other at all? Do different NRXNs — either different principal isoforms, or different splice variants — have distinct functions? α-NRXNs and β-NRXNs cannot be functionally redundant because the deletion of genes encoding α-NRXNs causes a massive phenotype — a phenotype that cannot be compensated for by the remaining β-NRXNs[10] — so what else do α-NRXNs do? Uncovering answers to these and many other questions will provide insight not only into the fundamental mechanisms of synaptic cell adhesion but also into the molecular determinants of neural-circuit properties. Moreover, the apparent involvement of NRXNs and NLGNs in different cognitive diseases begs the question of whether these diseases are distinct entities or form a continuum of mental dysfunctions. With the emerging findings on the genetics of cognitive diseases, a molecular nosology of cognitive diseases may become possible. Furthermore, if a participation of NRXNs and NLGNs in cognitive diseases is confirmed in more extensive studies, new diagnostic and therapeutic possibilities may emerge, for example by selectively modulating the NRXN–NLGN interaction. Again, much more work will be required to explore these possibilities, but the present results are encouraging in this direction as well. ■

1. Cowan, W. M., Südhof, T. C. & Stevens, C. F. (eds) *Synapses* (Johns Hopkins Univ. Press, 2000).
2. Rozov, A., Burnashev, N., Sakmann, B. & Neher, E. Transmitter release modulation by intracellular Ca^{2+} buffers in facilitating and depressing nerve terminals of pyramidal cells in layer 2/3 of the rat neocortex indicates a target cell-specific difference in presynaptic calcium dynamics. *J. Physiol. (Lond.)* **531,** 807–826 (2001).
3. Dityatev, A. & El-Husseini, A. (eds) *Molecular Mechanisms of Synaptogenesis* (Springer, 2006).
4. Abbott, L. F. & Regehr, W. G. Synaptic computation. *Nature* **431,** 796–803 (2004).
5. Linkenhoker, B. A., von der Ohe, C. G. & Knudsen, E. I. Anatomical traces of juvenile learning in the auditory system of adult barn owls. *Nature Neurosci.* **8,** 93–98 (2005).
6. Arikkath, J. & Reichardt, L. F. Cadherins and catenins at synapses: roles in synaptogenesis and synaptic plasticity. *Trends Neurosci.* doi:10.1016/j.tins.2008.07.001 (2008).
7. Salinas, P. C. & Zou, Y. Wnt signaling in neural circuit assembly. *Annu. Rev. Neurosci.* **31,** 339–358 (2008).
8. Craig, A. M. & Kang, Y. Neurexin–neuroligin signaling in synapse development. *Curr. Opin. Neurobiol.* **17,** 43–52 (2007).
9. Dean, C. & Dresbach, T. Neuroligins and neurexins: linking cell adhesion, synapse formation and cognitive function. *Trends Neurosci.* **29,** 21–29 (2006).
10. Missler, M. *et al.* α-Neurexins couple Ca^{2+} channels to synaptic vesicle exocytosis. *Nature* **423,** 939–948 (2003).
This paper shows that deletion of genes encoding α-NRXNs in mice causes a lethal presynaptic release phenotype and a loss of presynaptic Ca^{2+}-channel function.
11. Varoqueaux, F. *et al.* Neuroligins determine synapse maturation and function. *Neuron* **51,** 741–754 (2006).
This paper describes mice lacking NLGN1, NLGN2 and NLGN3, demonstrating that deletion of the genes encoding NLGNs is lethal because it results in impaired synaptic transmission and not a decrease in synapse numbers.
12. Chubykin, A. A. *et al.* Activity-dependent validation of excitatory vs. inhibitory synapses by neuroligin-1 vs. neuroligin-2. *Neuron* **54,** 919–931 (2007).
This paper demonstrates that the enhancement of synaptic transmission by overexpressed NLGN1 depends on NMDA-receptor signalling.
13. Feng, J. *et al.* High frequency of neurexin 1β signal peptide structural variants in patients with autism. *Neurosci. Lett.* **409,** 10–13 (2006).
14. Szatmari, P. *et al.* Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nature Genet.* **39,** 319–328 (2007).
15. Kim, H. G. *et al.* Disruption of neurexin 1 associated with autism spectrum disorder. *Am. J. Hum. Genet.* **82,** 199–207 (2008).
16. Yan, J. *et al.* Neurexin 1α structural variants associated with autism. *Neurosci. Lett.* doi:10.1016/j.neulet.2008.04.074 (2008).
17. Zahir, F. R. *et al.* A patient with vertebral, cognitive and behavioural abnormalities and a de novo deletion of NRXN1α. *J. Med. Genet.* **45,** 239–243 (2008).
18. Marshall, C. R. Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.* **82,** 477–488 (2008).
19. Kirov, G. Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. *Hum. Mol. Genet.* **17,** 458–465 (2008).
20. Walsh, T. *et al.* Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320,** 539–543 (2008).
21. Jamain, S. *et al.* Mutations of the X-linked genes encoding Nlgns NLGN3 and NLGN4 are associated with autism. *Nature Genet.* **34,** 27–29 (2003).
This paper describes the first mutations in the genes that encode NLGNs in patients with familial ASD, initiating a search for mutations in other families in NLGNs or their associated molecules.
22. Laumonnier, F. *et al.* X-linked mental retardation and autism are associated with a mutation in the *NLGN4* gene, a member of the neuroligin family. *Am. J. Hum. Genet.* **74,** 552–527 (2004).
23. Yan, J. *et al.* Analysis of the neuroligin 3 and 4 genes in autism and other neuropsychiatric patients. *Mol. Psychiatry* **10,** 329–332 (2005).
24. Talebizadeh, Z. *et al.* Novel splice isoforms for NLGN3 and NLGN4 with possible implications in autism. *J. Med. Genet.* **43,** e21 (2006).
25. Chocholska, S., Rossier, E., Barbi, G. & Kehrer-Sawatzki, H. Molecular cytogenetic analysis of a familial interstitial deletion Xp22.2–22.3 with a highly variable phenotype in female carriers. *Am. J. Med. Genet. A* **140,** 604–610 (2006).
26. Lawson-Yuen, A., Saldivar, J. S., Sommer, S. & Picker, J. Familial deletion within NLGN4 associated with autism and Tourette syndrome. *Eur. J. Hum. Genet.* **16,** 614–618 (2008).
27. Macarov, M. *et al.* Deletions of VCX-A and NLGN4: a variable phenotype including normal intellect. *J. Intellect. Disabil. Res.* **51,** 329–333 (2007).
28. Ushkaryov, Y. A., Rohou, A. & Sugita, S. α-Latrotoxin and its receptors. *Handb. Exp. Pharmacol.* **184,** 171–206 (2008).
29. Ushkaryov, Y. A., Petrenko, A. G., Geppert, M. & Südhof, T. C. Neurexins: synaptic cell surface proteins related to the α-latrotoxin receptor and laminin. *Science* **257,** 40–56 (1992).
This paper reports the discovery of NRXNs as presynaptic α-latrotoxin receptors.
30. Missler, M. & Südhof, T. C. Neurexins: three genes and 1001 products. *Trends Genet.* **14,** 20–25 (1998).
31. Peles, E. & Salzer, J. L. Molecular domains of myelinated axons. *Curr. Opin. Neurobiol.* **10,** 558–565 (2000).
32. Tabuchi, K. & Südhof, T. C. Structure and evolution of neurexin genes: insight into the mechanism of alternative splicing. *Genomics* **79,** 849–859 (2002).
33. Ullrich, B., Ushkaryov, Y. A. & Südhof, T. C. Cartography of neurexins: more than 1000 isoforms generated by alternative splicing and expressed in distinct subsets of neurons. *Neuron* **14,** 497–507 (1995).
34. Rozic-Kotliroff, G. & Zisapel, N. Ca^{2+}-dependent splicing of neurexin IIα. *Biochem. Biophys. Res. Commun.* **352,** 226–230 (2007).
35. Ushkaryov, Y. A. & Südhof, T. C. Neurexin IIIα: Extensive alternative splicing generates membrane-bound and soluble forms in a novel neurexin. *Proc. Natl Acad. Sci. USA* **90,** 6410–6414 (1993).
36. Sugita, S., Khvotchev, M. & Südhof, T. C. Neurexins are functional α-latrotoxin receptors. *Neuron* **22,** 489–496 (1999).
37. Berninghausen, O. *et al.* Neurexin Iβ and neuroligin are localized on opposite membranes in mature central synapses. *J. Neurochem.* **103,** 1855–1863 (2007).
38. Chubykin, A. A. *et al.* Dissection of synapse induction by neuroligins: effect of a neuroligin mutation associated with autism. *J. Biol. Chem.* **280,** 22365–22374 (2005).
39. Kattenstroth, G., Tantalaki, E., Südhof, T. C., Gottmann, K. & Missler, M. Postsynaptic *N*-methyl-D-aspartate receptor function requires α-neurexins. *Proc. Natl Acad. Sci. USA* **101,** 2607–2612 (2004).
40. Taniguchi, H. *et al.* Silencing of neuroligin function by postsynaptic neurexins. *J. Neurosci.* **27,** 2815–2824 (2007).
41. Ichtchenko, K. *et al.* Neuroligin 1: a splice-site specific ligand for α-neurexins. *Cell* **81,** 435–443 (1995).
This paper includes the identification of NLGNs as postsynaptic NRXN ligands.
42. Petrenko, A. G. *et al.* Structure and evolution of neurexophilin. *J. Neurosci.* **16,** 4360–4369 (1996).
43. Sugita, S. *et al.* A stoichiometric complex of neurexins and dystroglycan in brain. *J. Cell Biol.* **154,** 435–445 (2001).
44. Ichtchenko, K., Nguyen, T. & Südhof, T. C. Structures, alternative splicing, and neurexin binding of multiple neuroligins. *J. Biol. Chem.* **271,** 2676–2682 (1996).
45. Boucard, A., Chubykin, A. A., Comoletti, D., Taylor, P. & Südhof, T. C. A splice-code for trans-synaptic cell adhesion mediated by binding of neuroligin 1 to α- and β-neurexins. *Neuron* **48,** 229–236 (2005).
46. Rissone, A. *et al.* Comparative genome analysis of the neurexin gene family in *Danio rerio*: insights into their functions and evolution. *Mol. Biol. Evol.* **24,** 236–252 (2007).
47. Bolliger, M. F. *et al.* Unusually rapid evolution of neuroligin-4 in mice. *Proc. Natl Acad. Sci. USA* **105,** 6421–6426 (2008).
48. Song, J.-Y., Ichtchenko, K., Südhof, T. C. & Brose, N. Neuroligin 1 is a postsynaptic cell-adhesion molecule of excitatory synapses. *Proc. Natl Acad. Sci. USA* **96,** 1100–1125 (1999).
49. Varoqueaux, F., Jamain, S. & Brose, N. Neuroligin 2 is exclusively localized to inhibitory synapses. *Eur. J. Cell Biol.* **83,** 449–456 (2004).
50. Graf, E. R., Zhang, X., Jin, S. X., Linhoff, M. W. & Craig, A. M. Neurexins induce differentiation of GABA and glutamate postsynaptic specializations via neuroligins. *Cell* **119,** 1013–1026 (2004).
51. Budreck, E. C. & Scheiffele, P. Neuroligin-3 is a neuronal adhesion protein at GABAergic and glutamatergic synapses. *Eur. J. Neurosci.* **26,** 1738–1748 (2007).
52. Comoletti, D. *et al.* Gene selection, alternative splicing, and post-translational processing regulate neuroligin selectivity for β-neurexins. *Biochemistry* **45,** 12816–12827 (2006).
53. Chih, B., Gollan, L. & Scheiffele, P. Alternative splicing controls selective trans-synaptic interactions of the neuroligin–neurexin complex. *Neuron* **51,** 171–178 (2006).
54. Fabrichny, I. P. *et al.* Structural analysis of the synaptic protein neuroligin and its β-neurexin complex: determinants for folding and cell adhesion. *Neuron* **56,** 979–991 (2007).
References 54–56 report the first atomic structures of the neuroligin–neurexin complex.
55. Arac, D. *et al.* Structures of neuroligin-1 complex reveal specific protein–protein and protein–Ca^{2+} interactions. *Neuron* **56,** 992–1003 (2007).
56. Chen, X., Liu, H., Shim, A. H., Focia, P. J. & He, X. Structural basis for synaptic adhesion mediated by neuroligin–neurexin interactions. *Nature Struct. Mol. Biol.* **15,** 50–56 (2008).
57. Shen, K. C. *et al.* Regulation of neurexin 1β tertiary structure and ligand binding through alternative splicing. *Structure* **16,** 422–431 (2008).
58. Koehnke, J. *et al.* Crystal structures of β-neurexin 1 and β-neurexin 2 ectodomains and dynamics of splice insertion sequence 4. *Structure* **16,** 410–421 (2008).
59. Hata, Y., Butz, S. & Südhof, T. C. CASK: a novel dlg/PSD95 homologue with an N-terminal CaM kinase domain identified by interaction with neurexins. *J. Neurosci.* **16,** 2488–2494 (1996).
60. Biederer, T. & Südhof, T. C. CASK and protein 4.1 support F-actin nucleation on neurexins. *J. Biol. Chem.* **276,** 47869–47876 (2001).
61. Mukherjee, K. *et al.* CASK functions as a neurexin-kinase by an unusual mechanism. *Cell* **133,** 328–339 (2008).
62. Butz, S., Okamoto, M. & Südhof, T. C. A tripartite protein complex with the potential to couple synaptic vesicle exocytosis to cell adhesion in brain. *Cell* **94,** 773–782 (1998).
63. Borg, J. P. *et al.* Molecular analysis of the X11-mLin-2/CASK complex in brain. *J. Neurosci.* **19,** 1307–1316 (1999).
64. Atasoy, D. *et al.* Deletion of CASK in mice is lethal and impairs synaptic function. *Proc. Natl Acad. Sci. USA* **104,** 2525–2530 (2007).
65. Irie, M. *et al.* Binding of neuroligins to PSD-95. *Science* **277,** 1511–1515 (1997).
Together with the finding that NRXNs bind to the MAGUK CASK (ref. 48), this paper reveals a quasi-symmetrical design of the NRXN–NLGN junction that contains PSD95 bound to NLGNs postsynaptically.
66. Sheng, M. & Hoogenraad, C. C. The postsynaptic architecture of excitatory synapses: a more quantitative view. *Annu. Rev. Biochem.* **76,** 823–847 (2007).
67. Scheiffele, P., Fan, J., Choih, J., Fetter, R. & Serafini, T. Neuroligin expressed in nonneuronal cells triggers presynaptic development in contacting axons. *Cell* **101,** 657–669 (2000).
This paper describes the first evidence that NLGNs are not only localized to synapses but also function there, by showing that overexpressed NLGN1 or NLGN2 in a non-neuronal cell can induce co-cultured neurons to form synapses onto that cell.
68. Nam, C. I. & Chen, L. Postsynaptic assembly induced by neurexin–neuroligin interaction and neurotransmitter. *Proc. Natl Acad. Sci. USA* **102,** 6137–6142 (2005).
69. Chih, B., Afridi, S. K., Clark, L. & Scheiffele, P. Disorder-associated mutations lead to functional inactivation of neuroligins. *Hum. Mol. Genet.* **13,** 1471–1477 (2004).
70. Chih, B., Engelman, H. & Scheiffele, P. Control of excitatory and inhibitory synapse formation by neuroligins. *Science* **307,** 1324–1328 (2005).
71. Zhang, W. *et al.* Extracellular domains of α-neurexins participate in regulating synaptic transmission by selectively affecting N-and P/Q-type Ca^{2+}-channels. *J. Neurosci.* **25,** 4330–4342 (2005).
72. Dudanova, I. *et al.* Important contribution of α-neurexins to Ca^{2+}-triggered exocytosis of secretory granules. *J. Neurosci.* **26,** 10599–10613 (2006).
73. Lord, C., Cook, E. H., Leventhal, B. L. & Amaral, D. G. Autism spectrum disorders. *Neuron* **28,** 355–363 (2000).

74. Pardo, C. A. & Eberhart, C. G. The neurobiology of autism. *Brain Pathol.* **7,** 434–447 (2007).
75. Schmitz, C. & Rezaie, P. The neuropathology of autism: where do we stand? *Neuropathol. Appl. Neurobiol.* **34,** 4–11 (2008).
76. Courchesne, E. *et al*. Mapping early brain development in autism. *Neuron* **56,** 399–413 (2007).
77. Durand, C. M. *et al*. Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nature Genet.* **39,** 25–37 (2007).
78. Moessner, R. *et al*. Contribution of SHANK3 mutations to autism spectrum disorder. *Am. J. Hum. Genet.* **81,** 1289–1297 (2007).
79. Okamoto, N. *et al*. 22q13 Microduplication in two patients with common clinical manifestations: a recognizable syndrome? *Am. J. Med. Genet. A* **143A,** 2804–2809 (2007).
80. Manning, M. A. *et al*. Terminal 22q deletion syndrome: a newly recognized cause of speech and language disability in the autism spectrum. *Pediatrics* **114,** 451–457 (2004).
81. Jeffries, A. R. *et al*. Molecular and phenotypic characterization of ring chromosome 22. *Am. J. Med. Genet. A* **137,** 139–147 (2005).
82. Wilson, H. L. *et al*. Molecular characterisation of the 22q13 deletion syndrome supports the role of haploinsufficiency of SHANK3/PROSAP2 in the major neurological symptoms. *J. Med. Genet.* **40,** 575–584 (2003).
83. Tobaben, S., Südhof, T. C. & Stahl, B. The G protein-coupled receptor CL1 interacts directly with proteins of the Shank family. *J. Biol. Chem.* **275,** 36204–36210 (2000).
84. Morrow, E. M. *et al*. Identifying autism loci and genes by tracing recent shared ancestry. *Science* **321,** 218–223 (2008).
85. Comoletti, D. *et al.* The Arg451Cys-neuroligin-3 mutation associated with autism reveals a defect in protein processing. *J. Neurosci.* **24,** 4889–4893 (2004).
86. Tabuchi, K. *et al*. A neuroligin-3 mutation implicated in autism increases inhibitory synaptic transmission in mice. *Science* **318,** 71–76 (2007).
This paper describes a mouse model of ASD in which a point mutation found in two brothers with ASD (Arg451Cys in NLGN3) was introduced into mice by homologous recombination.
87. Jamain, S. *et al*. Reduced social interaction and ultrasonic communication in a mouse model of monogenic heritable autism. *Proc. Natl Acad. Sci. USA* **105,** 1710–1715 (2008).
88. Hishimoto, A. *et al*. Neurexin 3 polymorphisms are associated with alcohol dependence and altered expression of specific isoforms. *Hum. Mol. Genet.* **16,** 2880–2891 (2007).
89. Lachman, H. M. *et al.* Genomewide suggestive linkage of opioid dependence to chromosome 14q. *Hum. Mol. Genet.* **16,** 1327–1334 (2007).
90. Meinrenken, C. J., Borst, J. G. & Sakmann, B. Local routes revisited: the space and time dependence of the Ca^{2+} signal for phasic transmitter release at the rat calyx of Held. *J. Physiol. (Lond.)* **547,** 665–689 (2003).
91. Südhof, T. C. The synaptic vesicle cycle. *Annu. Rev. Neurosci.* **27,** 509–547 (2004).

Acknowledgements I thank D. Arac, A. Brunger, X. Liu, J. Gibson and K. Huber for advice and help with figures. Work in my laboratory on NRXNs and NLGNs is supported by the National Institute of Mental Health and the Simon Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Correspondence should be addressed to the author (tcs1@stanford.edu).

NATURE|Vol 455|16 October 2008|doi:10.1038/nature07457

Failure of neuronal homeostasis results in common neuropsychiatric phenotypes

Melissa B. Ramocki[1] & Huda Y. Zoghbi[1,2]

Failure of normal brain development leads to mental retardation or autism in about 3% of children. Many genes integral to pathways by which synaptic modification and the remodelling of neuronal networks mediate cognitive and social development have been identified, usually through loss of function. Evidence is accumulating, however, that either loss or gain of molecular functions can be deleterious to the nervous system. Copy-number variation, regulation of gene expression by non-coding RNAs and epigenetic changes are all mechanisms by which altered gene dosage can cause the failure of neuronal homeostasis.

Homeostasis is the ability of a system to return to a set point following perturbation. Much is understood about neuronal homeostatic mechanisms at the level of physiological output (for example, the nerve action potential), but the underlying molecular processes are the subject of intense investigation. The developing and learning brain is composed of neurons that must continue to carry out routine functions even while neurogenesis is occurring and the synaptic connections between neurons are being strengthened, remodelled and pruned. Neurons within affected networks must therefore make homeostatic responses to re-establish a proper balance of excitation and inhibition. Recent clinical studies show that loss or gain in dosage of proteins or RNAs in different neurodevelopmental syndromes often results in similar or overlapping sets of neurological symptoms, suggesting that the molecules involved in cognitive and behavioural processes participate in highly regulated homeostatic mechanisms. This inability of neurons *in vivo* to compensate for seemingly minor alterations in protein or RNA dosage should teach us something about the nervous system. It may help us to answer questions such as: Why are there hundreds of known genes whose alteration causes mental retardation or autism, yet rarely can we predict genotype solely on the basis of neurophenotype? And why is autism a prominent feature of so many disorders affecting cognition?

Inherent in this discussion is the fact that single genes or proteins do not encode specific behaviours or neurological deficits but instead encode biological functions. When disruption of biological processes leads to neuronal dysfunction in a developing brain, numerous phenotypes may result, depending on the function(s) subserved by the affected cell type and other neurons within the network. In other words, neurodevelopmental disorders are not cell autonomous because even neuron-specific changes will always affect the neuronal network. Loss or gain of function of any protein that alters synaptic output may cause neurological or psychiatric phenotypes because changes in neuronal excitability demand compensation, which may, in turn, exhaust the homeostatic capacity of the network. *In vitro* and *ex vivo* model systems have demonstrated the exceptional ability of neurons to compensate for experimental perturbations by modulating ion channels, receptors, signalling pathways and neurotransmitters. At the molecular level, such homeostatic processes require chromatin remodelling, changes in gene expression and repression, changes in protein production and turnover, and cytoskeletal rearrangement.

The disorders discussed in this Review all represent the cumulative impact of inappropriate gene dosage on a neuronal system and variously involve abnormal regulation of gene expression by small non-coding RNA molecules, alteration of RNA metabolism leading to abnormal protein synthesis, abnormal protein turnover, abnormal chromatin modulation and gene expression, and altered cytoskeletal dynamics. Amazingly, these disorders show that the inability to maintain neuronal homeostasis at the level of a variety of molecular processes is sufficient to cause common end points such as mental retardation, epilepsy and autism (Fig. 1). There are almost no medical therapies for these conditions, yet we know from animal models of some of them — for example, Angelman syndrome, fragile X syndrome and Rett syndrome — that the therapeutic potential exists. Thus, never before has the promise for neuroscience to uncover the molecular underpinnings of cognition, social interaction and behaviour been greater, nor has the potential to discover treatments that restore homeostasis to a diseased brain been riper.

In this Review, we discuss the pathophysiology of six neurodevelopmental disorders that illustrate instances in which loss or gain of molecular function converge on the common phenotypes of mental retardation and autism spectrum behaviours. We put forward a hypothesis in which the failure of neuronal homeostasis or homeostatic compensation leads to neuronal networks with weakened synaptic flexibility and to common clinical end points.

Abnormal protein synthesis disrupts synaptic function

Levels of RNA and protein are regulated by transcription and translation, respectively. Whether a messenger RNA molecule will be translated into protein often depends on the presence of RNA-binding proteins that regulate the translation of specific mRNA molecules, thus providing a mechanism for regulating temporal and spatial patterns of protein expression. Fragile X syndrome, the disorder we discuss first, involves abnormal dosage of an RNA-binding protein that results in a devastating neurological outcome.

The discovery of a CGG repeat expansion in the 5′ untranslated region of the gene *FMR1* led to insight into the pathophysiology of a common form of heritable mental retardation, fragile X syndrome[1]. Hypermethylation of the expanded CGG repeats and upstream CpG island leads to transcriptional silencing of *FMR1* (ref. 2). Mental retardation, characteristic facies (facial appearance), macro-orchidism, autism and

[1]Department of Pediatrics, Section of Pediatric Neurology and Developmental Neuroscience, Baylor College of Medicine, 1 Baylor Plaza, MS 225, BCMT-T807, Houston, Texas 77030, USA. [2]Department of Pediatrics, Section of Human and Molecular Genetics, Baylor College of Medicine, 1 Baylor Plaza, MS 225, BCMT-T807, Houston, Texas 77030, USA.

epilepsy occur with a full mutation of >200 repeats[2]. *Fmr1*-null mice have abnormal maturation of, and pruning of, dendritic spines, resulting in abnormal dendritic-spine morphology similar to that observed in the cerebral cortex of patients with fragile X syndrome[3] (Fig. 2b). Transgenic mice that overexpress a human *FMR1* transgene develop hypoactivity, anxiety and abnormal responses to sensory stimulation[4] and predict a human gain-of-function disorder.

The product of normal *FMR1* is an RNA-binding protein, FMRP, that associates with polyribosomes and suppresses the translation of specific mRNA molecules[2]. Identification of mRNAs differentially associated with the FMRP ribonucleoprotein complex suggested that activity-dependent alterations in local protein synthesis at synapses leads to the neurological consequences of fragile X syndrome[5,6]. The fact that activation of group 1 metabotropic glutamate receptors (mGluR) stimulates synaptic protein synthesis in an mRNA-translation-dependent manner led to the mGluR theory of fragile X pathogenesis[7,8]. In the absence of FMRP, mGluR5 signalling stimulates increased internalization of the ionotropic AMPA receptors for glutamate[9]. Also, the mGluR-induced translation of proteins that are crucial for surface expression of the AMPA receptor, such as AMPA receptor subunits and postsynaptic-density protein 95 (PSD95), does not occur[10]. These data support a model in which loss of FMRP leads to an increase in baseline protein synthesis to the extent that synaptic activation does not induce the additional local protein synthesis necessary to regulate synaptic plasticity appropriately[10].

Treatment with a non-competitive mGluR5 antagonist rescues the AMPA-trafficking defect in cultured hippocampal neurons[9], the seizure phenotype of *Fmr1*-null mice[11], and the synaptic plasticity, courtship behaviour and mushroom-body defects in a *Drosophila* model of fragile X syndrome[12]. Furthermore, a 50% reduction in mGluR5 expression in *Fmr1*-null mice rescued synaptic plasticity, dendritic-spine density, protein synthesis, and memory and epilepsy phenotypes[8]. This discovery, and the fact that a 50% reduction in mGluR5 does not lead to overt behavioural phenotypes in normal mice[8], has generated much excitement over the potential of mGluR5 antagonists to improve the neurodevelopmental phenotypes of fragile X syndrome. In summary, FMRP and mGluR5 signalling seem to oppose each other functionally in the regulation of local synaptic protein synthesis in an activity-dependent, homeostatic manner. In addition to causing abnormal protein synthesis and dynamically weakened neurons (neurons with attenuated ability to respond to change), loss of FMRP probably also causes secondary homeostatic changes that further damage neuronal networks.

MicroRNAs and cognition

Small non-coding microRNAs (miRNAs) that repress the translation of target mRNAs are emerging as important second-tier regulators of protein levels. In our next example, that of the heterozygous 22q11.2 deletion and duplication syndromes, altered dosage of an RNA-binding protein important for the processing of miRNAs may underlie significant neurodevelopmental and psychiatric morbidity.

The 22q11.2 deletion syndrome is caused by the loss of the q11.2 region from one copy of chromosome 22. Individuals with this deletion syndrome have highly variable phenotypes, including the phenotypic constellations that are typically associated with DiGeorge syndrome

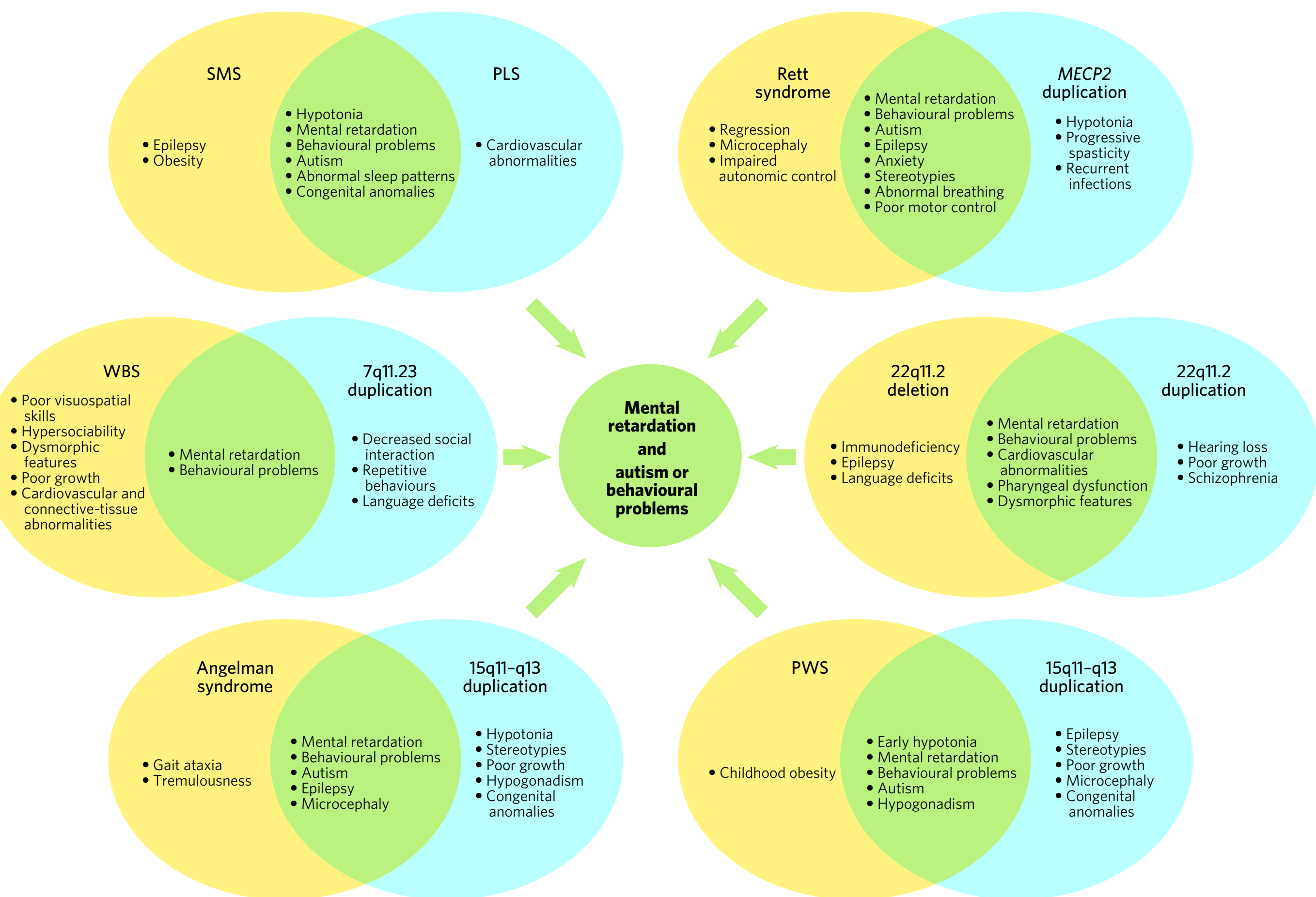


Figure 1 | Loss of protein or RNA function causes neurodevelopmental disorders with phenotypes overlapping those caused by gain of protein or RNA function. Each pair of ovals demonstrates loss and duplication of the same chromosomal region. Phenotypes unique to loss of function are shown in yellow; phenotypes unique to gain of function are shown in blue; phenotypes common to both loss of function and gain of function are shown in green. Each of these alterations in neuronal function converges on pathways that cause mental retardation and autism or abnormal behaviours. (Both Prader–Willi syndrome (PWS) and Angelman syndrome (bottom) involve deletion of 15q11–q13.) *MECP2*, gene encoding methyl-CpG-binding protein 2; PLS, Potocki–Lupski syndrome; SMS, Smith–Magenis syndrome; WBS, Williams–Beuren syndrome.

and/or velocardiofacial syndrome. Developmental and neuropsychiatric phenotypes are prevalent in individuals with the 22q11.2 deletion syndrome and include developmental delay, learning difficulties, epilepsy, hyperactivity, anxiety, autism, obsessive–compulsive disorder, bipolar disorder and schizophrenia[13–15]. In fact, 22q11.2 deletion conveys the strongest DNA-based risk factor identified so far for developing schizophrenia[14]. Unexpectedly, the duplication of 22q11.2 (the reciprocal of the 22q11.2 deletion) was recently identified in individuals with developmental delay, learning disabilities, behavioural problems, cardiovascular abnormalities, velopharyngeal insufficiency, growth retardation and dysmorphic features — phenotypes similar to those of individuals with the 22q11.2 deletion syndrome[16,17] (Fig. 1).

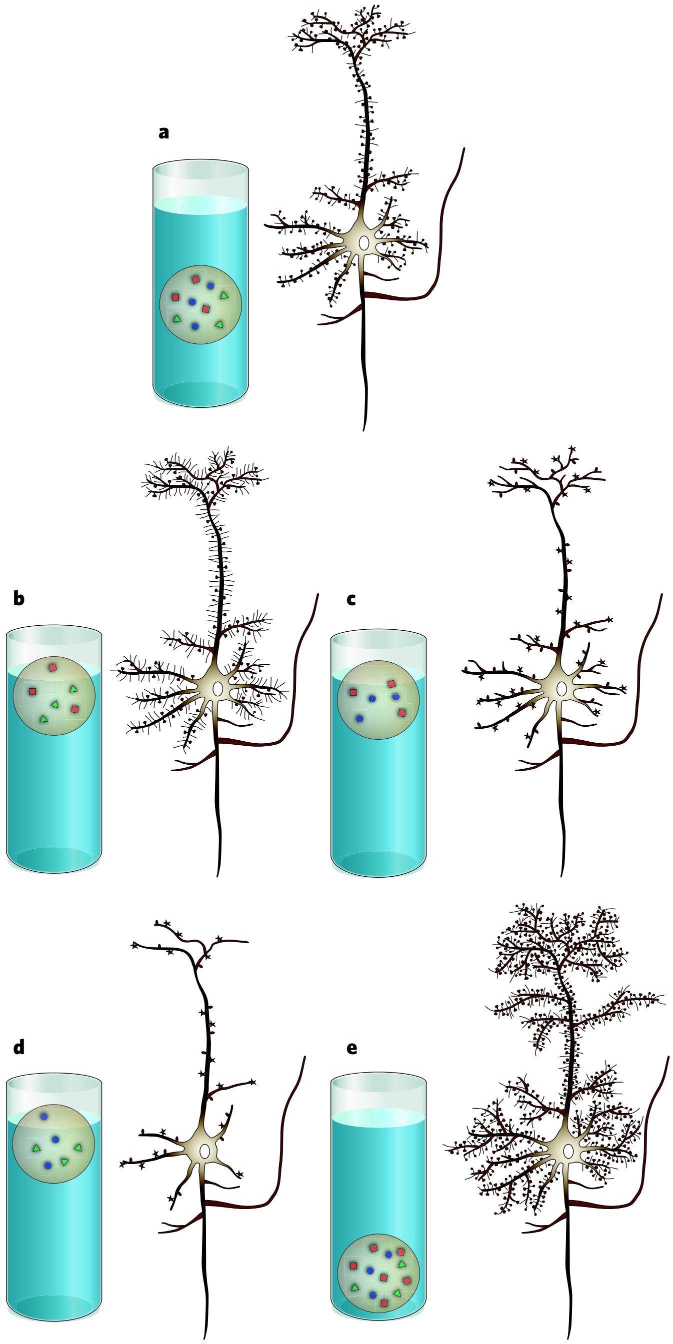


Figure 2 | Loss or gain of protein or RNA function results in altered neuronal homeostasis or 'imbalance'. Homeostasis is represented by the balance of a weighted ball floating in water. Individual proteins involved in the following syndromes are each represented by a coloured shape. **a**, Representation of a healthy neuron, with proteins in balance. **b**, Representation of a neuron in fragile X syndrome, with normal dendrites but an increased density of longer, thinner, immature-appearing dendritic spines. **c**, Representation of a neuron in Angelman syndrome, with normal dendrites but a decreased density of, and abnormally shaped, dendritic spines. **d**, Representation of a neuron in Rett syndrome or *Mecp2*-null mice, with decreased density of dendrites and decreased density of, and abnormally shaped, synapses (represented by dendritic spines). **e**, Representation of a neuron from mice in which *Mecp2* is duplicated, with increased density of synapses (represented by increased density of dendrites and dendritic spines).

Cardiovascular anomalies and impaired sensorimotor gating (the abnormal modulation of sensory information transmitted to a motor system, resulting in abnormal behavioural output) are reproduced in a mouse model with deletion of the region homologous to human 22q11.2 (refs 18, 19). Mutation of a single gene in the critical region, *TBX1* (which encodes a T-box-containing transcription factor), reproduces most physical features of DiGeorge syndrome and velocardiofacial syndrome in human patients[20]. *TBX1* mutation has been implicated in nonspecific cognitive dysfunction and in a boy with Asperger's syndrome[18], and haploinsufficiency for *Tbx1* or *Gnb1l* (which encodes a G-protein-β-subunit-like protein) is sufficient to cause abnormal sensorimotor gating in mice[18]. Transgenic mice that overexpressed a region of human 22q11.2 encompassing four genes, *CDCREL* (which encodes a substrate for the ubiquitin–protein ligase parkin), *GP1B* (which encodes a platelet-surface membrane protein), *TBX1* and *GNB1L*, showed hyperactivity and a lack of habituation that could be ameliorated by treatment with antipsychotic medication[21]. Other genes in the 22q11.2 deletion region have been put forward as candidate genes for schizophrenia; however, none has been incontrovertibly linked to psychiatric or neurodevelopmental phenotypes[22].

Transcriptional profiling in a mouse model lacking the 22q11.2-homologous region revealed consistent upregulation of miRNA-related transcripts in the hippocampus and prefrontal cortex[23]. One of the genes mapping to the 22q11.2 deletion region is *DGCR8*, which encodes a double-stranded RNA-binding protein that is part of a complex that processes primary miRNAs to mature miRNAs. Thus Kimberly Stark *et al.*[23] proposed that DGCR8 might mediate this effect. Indeed, haploinsufficiency for *Dgcr8* in mice resulted in abnormal miRNA biogenesis, smaller dendritic spines, a less complex dendritic tree and impaired sensorimotor gating and spatial working memory, suggesting the hypothesis that abnormal processing of miRNAs may lead to the neurological and behavioural phenotypes observed in the 22q11.2 deletion syndrome[23], perhaps through the inability of neurons to regulate translational responses to stimuli. Like the loss of FMRP, the loss of DGCR8 may lead to dynamically weakened neurons and compensatory homeostatic mechanisms that damage neuronal networks. The study by Stark and colleagues[23] put forward a new role for miRNAs in cognitive processes and predicted that other components of this pathway might contribute to neurological and psychiatric disease in humans.

Imprinting failure disrupts protein and RNA dynamics

A key mechanism for fine-tuning gene dosage is genomic imprinting, the phenomenon in which only the maternal or paternal copies of certain genes are expressed as a result of differential cytosine methylation on the maternally and paternally inherited homologous chromosomes. Biological processes that are extremely vulnerable to altered protein dosage must exist to explain why some gene dosages are so critical that imprinting is necessary. The disorders discussed next illustrate the extreme dosage sensitivity of certain genes required for proper neuronal function.

Chromosome region 15q11–q13 is the best-characterized imprinted region in the human genome[24] because altered gene dosage resulting from uniparental disomy, imprinting-centre mutations or hemizygous deletion results in two distinct disorders: Angelman syndrome (which involves the loss of maternal genetic information) and Prader–Willi syndrome (PWS, which involves the loss of paternal genetic information)[25]. Individuals with Angelman syndrome typically show mental retardation, autism, microcephaly, gait ataxia, tremor, epilepsy, and inappropriate hand flapping, laughing and excitability[24]. PWS is characterized by infantile hypotonia (abnormally low muscle tone) and failure to thrive, which progresses to excessive eating and obesity by early childhood, and mental

retardation, hypogonadism and abnormal behaviours (including autism, psychosis and obsessive–compulsive behaviour)[26]. An overlapping clinical syndrome (Fig. 1) caused by increased maternal copy number of 15q11–q13 (the critical region for Angelman syndrome and PWS) also occurs, and includes hypotonia, mental retardation, autism, stereotypies, epilepsy and a spectrum of congenital anomalies[25].

The gene responsible for Angelman syndrome has been identified as that encoding ubiquitin–protein ligase E3A (UBE3A), because point mutations in this gene are sufficient to cause Angelman syndrome[24]. UBE3A marks proteins for degradation by the proteasome; however, the target proteins that fail to be degraded in Angelman syndrome have not been identified[24]. UBE3A is expressed from the maternal allele in the brain but is expressed from both alleles in peripheral tissues, suggesting that, to maintain proper function, neurons in the brain have a dosage requirement that differs from that of other cell types[24]. Mice deficient in the maternal *Ube3a* allele have abnormal dendritic spines (Fig. 2c), microcephaly, impaired motor function, impaired spatial learning and epilepsy, closely mimicking the human disorder[27]. The recent identification of an individual with typical PWS and paternal deletion of the genes encoding small nucleolar RNAs (snoRNAs, which are non-coding RNA molecules generally involved in RNA processing) confirmed previous hypotheses that paternal deficiency of the HBII-85 snoRNA cluster is sufficient to cause most features of PWS[26,28]. Mouse models with deletion of the homologous snoRNA cluster reproduce the neonatal hypotonia, growth retardation, abnormal motor learning, increased anxiety and hyperphagia observed with larger deletions[29].

It follows that the identification of UBE3A and HBII-85 targets is likely to demonstrate that failure of normal protein turnover in Angelman syndrome and failure of proper RNA processing in PWS also lead to dynamically weakened neurons and neuronal networks that are at the mercy of secondary compensatory mechanisms. The identification of these target proteins and RNAs will pinpoint novel pathways that could be targeted therapeutically.

A dosage-sensitive transcription factor

It is no surprise that improper dosage of a transcription factor essential for the regulation of gene expression can lead to detrimental consequences for neuronal networks in a developing brain. The disorders discussed in this section are caused by the altered dosage of such a protein.

Smith–Magenis syndrome (SMS), a constellation of infantile hypotonia and failure to thrive, congenital anomalies, mental retardation, behavioural problems (including self-mutilation, aggression, hyperactivity, stereotypies and often autism), obesity, sleep disturbance, decreased pain sensitivity and epilepsy[30,31], is in most individuals due to heterozygous deletion of 3.7 Mb in chromosome 17p11.2 (refs 32, 33). A few patients are haploinsufficient for the retinoic-acid-induced 1 (*RAI1*) gene product as a result of truncating mutations, indicating that *RAI1* is the primary dosage-sensitive culprit in SMS[30,34].

In a mouse cell line (P19 embryonal carcinoma cells), *Rai1* is upregulated subsequent to retinoic-acid-induced neuronal differentiation[35]. The RAI1 protein contains a bipartite nuclear-localization signal, polyserine and polymorphic polyglutamine tracts, and a carboxy-terminal PHD/zinc-finger domain[36]. The amino-terminal region of RAI1 is sufficient to activate transcription in heterologous reporter gene assays[37]. Therefore, RAI1 may function in chromatin remodelling and transcriptional processes[31]. Targeted disruption of *Rai1* in mice led to embryonic or perinatal lethality and severe craniofacial and skeletal abnormalities in the homozygous state[37], although a small percentage of null animals survived the neonatal period and frequently developed epilepsy and motor and learning deficits[38]. *Rai1*-haploinsufficient mice experienced obesity, less severe craniofacial abnormalities[37] and less frequent epilepsy[38].

Duplications of 17p11.2 (the reciprocal of the SMS-causing microdeletion) cause Potocki–Lupski syndrome (PLS), which is characterized by infantile hypotonia and failure to thrive, mental retardation, expressive language deficits, autism, sleep-disordered breathing and cardiovascular abnormalities[39,40]. The duplication-critical region has been narrowed down to a 1.3-Mb interval containing the *RAI1* locus; however, contributions of other genes or regulatory regions to the PLS phenotype cannot be excluded[40]. Mice with duplication of the region homologous to human chromosome 17p11.2 (located in mouse chromosome 11) are abnormally small and hyperactive and have deficits in learning and memory[41]. Transgenic mice that overexpress *Rai1* show growth retardation, hyperactivity, anxiety and increased sensitivity to pain, as well as impaired gait, forelimb grip strength and sensorimotor activity; these phenotypes worsen as gene dosage increases[31].

Altered *RAI1* dosage may not explain all of the phenotypes associated with SMS and PLS, but *RAI1* seems to be the most dosage-sensitive gene in the 17p11.2 deletion/duplication region. Gain or loss of RAI1 function leads to overlapping neurological syndromes (Fig. 1). Individuals with SMS can interact socially to some extent but have severely maladaptive and aggressive behaviour[30], and individuals with PLS primarily develop autistic features[40], suggesting that *RAI1* dosage is crucial in pathways that modulate these behaviours. Because little is known about the function and targets of RAI1, the next logical steps to delineating the underlying molecular pathways are to carry out expression array studies and genome-wide binding-site analysis of brain tissue from animals with loss or gain of RAI1 protein dosage, as well as to identify the binding partners of RAI1. Such studies should also provide exciting insight into the role of RAI1 in cognitive and behavioural processes.

Chromatin modulation essential for synaptic homeostasis

Chromatin remodelling occurs through mechanisms such as modification of histones, methylation of cytosine, regulation of DNA-binding proteins and recruitment of transcriptional activator and repressor protein complexes. The disorders discussed in this section are the result of improper dosage of a prototypical chromatin-remodelling protein.

Rett syndrome is one of the most common causes of mental retardation in females[42]. After 6 to 18 months of apparently normal development, affected females lose speech, social-interaction skills and the ability to learn[43]. They develop characteristic hand stereotypies, anxiety, microcephaly, epileptic seizures and impaired motor and autonomic control[43]. Mutations in the X-linked gene encoding methyl-CpG-binding protein 2 (*MECP2*) cause Rett syndrome[44]. Recently, males with Xq28 duplications that include the *MECP2* locus were found with infantile hypotonia, mental retardation, poor speech development, recurrent infections, epilepsy, progressive spasticity and autistic behaviours[45–49], features that overlap with the Rett syndrome phenotype (Fig. 1). The smallest region of overlap among the duplications studied suggests that *MECP2* is the primary dosage-sensitive gene in the duplicated region and is responsible for most of the observed phenotypes[45–49].

Mouse models of altered *Mecp2* gene dosage exist and were reviewed recently[43]. Briefly, male mice that produce no MeCP2 protein are normal at birth but develop tremor, gait abnormalities, reduced spontaneous movements, and weight and breathing abnormalities, and they die by 10–12 weeks of age[50]. Transgenic models in which mice overexpress human *MECP2* also exist, including *MECP2*^{Tg1} mice (which produce twice as much MeCP2 as wild-type controls) and *MECP2*^{Tg3} mice (which produce three times as much MeCP2 as wild-type mice). *MECP2*^{Tg1} males are normal until 10–12 weeks of age, when they develop a progressive neurological disorder that includes a forepaw stereotypy, seizures, aggression, hypoactivity, spasticity, kyphosis (spinal curvature), poor grooming and ataxia; they die by 12 months of age[51]. *MECP2*^{Tg3} males model the human *MECP2* triplication disorder[45], are smaller than their wild-type littermates at birth, develop the most severe progressive neurological disorder, and die between 3 and 6 weeks of age[51]. Recent data from a *Mecp2*$^{Flox/Y}$ mouse model, in which mice express 50% less *Mecp2* mRNA, and approximately 40% less protein, than wild-type mice, suggest that even this reduction in MeCP2 dosage is sufficient to cause neurological phenotypes, and predict that mild reductions in MeCP2 protein function might cause neuropsychiatric phenotypes in humans[52].

The fact that loss of MeCP2 function leads to neurological syndromes that overlap those caused by gain of MeCP2 function suggests that MeCP2 modulates crucial neuronal functions. Roles for MeCP2 in histone-deacetylase-dependent transcriptional repression, assembly

of secondary chromatin structure and regulation of RNA splicing have been demonstrated, but how these functions contribute to disease is not well understood[51]. Recent expression data from the hypothalami of *Mecp2*-null and *MECP2*^{Tg3} mice show that loss of MeCP2 function has an effect on gene expression that is opposite to that of gain of function, with MeCP2 gain primarily leading to transcriptional activation[53]. This study provided *in vivo* confirmation of an earlier genome-wide analysis that correlated MeCP2-binding sites with activation of nearby genes in a neuroblastoma cell line[54]. Whether this activation is direct or indirect is not yet known. Hippocampal glutamate-utilizing neurons that lack MeCP2 have a 46% reduction in synaptic response, and neurons producing twice the normal amount of MeCP2 have a twofold increase in synaptic response[55]. These observations were explained by concomitant changes in the formation of glutamate-utilizing synapses[55].

MeCP2 is phosphorylated on the serine residue at position 421 in response to neuronal activity, and this modification is necessary for MeCP2 to modulate dendritic growth and spine maturation[56]. Furthermore, MeCP2 regulates the expression of the gene encoding brain-derived neurotrophic factor (*Bdnf*), an activity-dependent gene[53,56,57]. *Mecp2*$^{308/Y}$ mice (which have an early truncation of the protein at amino acid 308) model human Rett syndrome and have heightened physiological responses to stress that are manifested by increased anxiety and increased levels of corticosterone[58]. The gene encoding corticotropin-releasing hormone (*Crh*) is a direct target of MeCP2, and altered *Crh* expression occurs only in those brain regions where *Crh* is normally expressed[58].

Collectively, these findings demonstrate that MeCP2 modulates gene expression in neurons in response to changing physiological states, underscore its important role in neuronal plasticity, and indicate that disorders resulting from MeCP2 loss of function or gain of function are caused by a failure of synaptic homeostasis (Fig. 2d, e).

Cytoskeletal modulation and molecular homeostasis

Dendrites and dendritic spines are crucial for synaptic function. For example, modulation of the cytoskeleton must occur during development of the nervous system and throughout life to regulate the trafficking and anchoring of neurotransmitter receptors in the postsynaptic density. The list of genes that are crucial for the regulation of dendritic spines and synapse structure, and that are altered in human mental retardation and autism syndromes, is growing and now includes *MECP2*, *FMR1*, *UBE3A*, *SHANK3* (which encodes a synaptic scaffolding protein), *NLGN3* (neuroligin 3) and *NLGN4*, and *NRXN1* (neurexin 1)[59]. Other genes that contribute to the development and modulation of synapses are those important to regulation of the cytoskeleton. The disorders discussed next are examples of how altered protein dosage leads to failure of cytoskeletal dynamics and dysfunctional neuronal systems.

Williams–Beuren syndrome (WBS) is caused by a hemizygous deletion of around 25 genes at 7q11.23 (ref. 60). Typical features include learning problems, mental retardation, visuospatial deficits, relative sparing of expressive language, inappropriate hypersociability, attention problems, dysmorphic facies, and cardiovascular and connective-tissue abnormalities[60]. Reciprocal duplication of this region causes language delay, decreased social interaction and repetitive behaviours[60,61]. Epilepsy and abnormal cortical migration have also been observed[62].

The cardiovascular phenotype that is commonly seen in WBS, supravalvular aortic stenosis, is caused by haploinsufficiency of the elastin gene[63]. Deletion of *CYLN2* (which encodes cytoplasmic linker protein 115, CLIP115), *LIMK1* (a LIM kinase) and *FZD9* (the receptor frizzled 9) may contribute to the cognitive phenotypes observed in WBS[64]. Frizzled 9 is a WNT receptor that is selectively expressed in the hippocampus[65]. *Fzd9*-hemizygous mice show increased apoptosis in the developing hippocampal dentate gyrus, a decreased seizure threshold, and impaired visuospatial learning[65]. Individuals with WBS have overall smaller brain volumes with decreased subcortical white matter, increased cortical folding and abnormal cell-packing densities in some regions, raising the possibility that *FZD9* contributes to such phenotypes[64].

In mice, *Cyln2* is expressed by many neurons and encodes the highly conserved protein CLIP115, which specifically associates with the ends of growing microtubules to regulate microtubule dynamics[64]. *Cyln2* hemizygosity leads to growth deficiency, altered brain morphology, altered hippocampal synaptic plasticity and specific motor coordination deficits in mice[66]. Microtubule dynamics are not altered, but increased accumulation of CLIP170 (another cytoplasmic linker protein) and dynactin occurs at microtubule ends in fibroblasts cultured from these animals[66]. More work is necessary to understand the contribution of the CLIP115 protein to cognitive function.

LIMK1 encodes a serine/threonine kinase that is expressed primarily in the central nervous system[67]. LIMK1 functions downstream of RHO signalling to regulate actin dynamics through inactivation of the actin-depolymerization factor cofilin by phosphorylation[67]. Brain tissue from *Limk1*-knockout mice showed decreased cofilin phosphorylation, decreased accumulation of actin filaments in dendritic spines, abnormal clusters of actin filaments in dendrites, altered spine morphology, and altered presynaptic and postsynaptic function (including enhanced hippocampal long-term potentiation)[67]. In regard to behaviour, the gene-knockout mice showed abnormal conditioned and spatial learning[67]. The importance of LIMK1 to spine morphogenesis and synaptic function underscores the role of synaptic homeostasis in neurological disorders[67–69].

These results suggest that loss or gain of CLIP115 or LIMK1 results in neurons with decreased ability to dynamically regulate cytoskeletal structure in response to the requirements of synaptic strengthening, remodelling and pruning processes.

Homeostatic control mechanisms control network function

The overlap of neurodevelopmental and psychiatric phenotypes (such as mental retardation, epilepsy, autism and other abnormal behaviours) that results from either loss or gain of the same proteins or RNA molecules supports an emerging theme that normal cognition and behaviour depend on tight neuronal homeostatic control mechanisms. Neurons must constantly respond to changes in their excitability that depend on the stage of development and environmental factors. This plasticity can be affected by a variety of molecular processes that influence synaptic output, including chromatin state, RNA biogenesis, transcription, translation, protein turnover and cytoskeletal dynamics.

Why is it that neurons *in vivo* are less able than neurons *in vitro* to compensate for what seem to be minor alterations in gene dosage? One possibility is that neuronal functions that are crucial for cognition, language and social development are particularly vulnerable. The extent of phenotypic overlap in these areas, irrespective of the molecular or cellular deficits, is curious. This is not to say that the disorders or phenotypes discussed here are homogeneous; rather, overlapping cognitive, behavioural and social domains are commonly affected. Such overlap is not observed with other neurological signs, such as neuromuscular weakness or dystonia. Another possibility is that compensatory changes within individual neurons are insufficient to restore normal activity to a network. We suggest a model in which altered proteins that are themselves integral to neuronal homeostasis (for example, FMRP or MeCP2), or loss or gain of function of proteins that trigger homeostatic responses, lead to dysfunctional neuronal networks with weakened synaptic flexibility (Fig. 3). We further propose that this lack of synaptic flexibility is what eventually causes the overlapping phenotypes.

The ability of a neuronal network to be flexible is limited by its inhibitory and excitatory connectivities, the intrinsic variability of cells, and its developmental stage and epigenetic history (acquired changes in patterns of gene expression that are not due to alterations in DNA sequence). A plausible explanation for periods of normal development before disease onset in neurodevelopmental or psychiatric disorders is that cells are able to compensate within a network for periods of time before an accumulation of variable and stochastic events causes system dysfunction. For example, it is interesting that girls with Rett syndrome have a period of normal development but that boys with the *MECP2*-duplication syndrome do not. Loss of MeCP2 function ultimately leads to clinical syndromes that overlap with those caused by gain of MeCP2 function, despite these changes in MeCP2 resulting in opposite molecular signatures

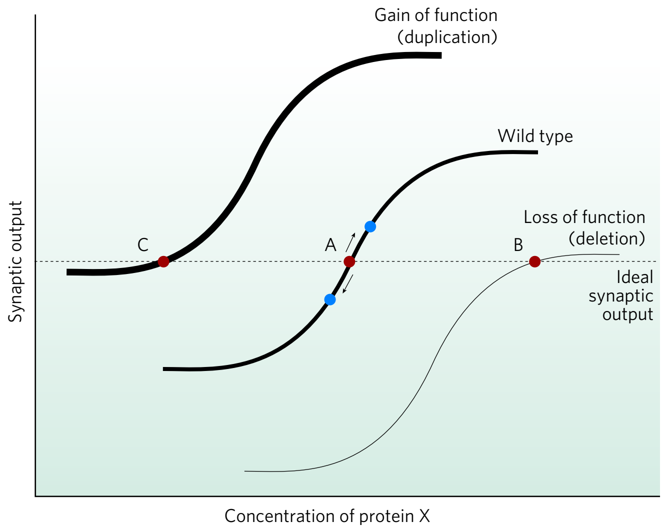


Figure 3 | Homeostatic responses could result in a compensated neuronal network with decreased flexibility. The curves show how synaptic output varies with the concentration of a given protein (X). Synaptic output is the product of synapse number and the synaptic strength per individual synapse. Protein X represents any protein important to synaptic strength and learning, for example, the AMPA receptor. The horizontal dashed line represents the ideal level of synaptic output (red circles). The middle curve (wild type) represents the normal situation. Point A is the concentration of X necessary to produce an ideal synaptic output. The neuron is flexible; in response to any learning event or stimulus, it can rapidly increase or decrease synaptic output (blue circles) with minimal change in the concentration of protein X (because the neuron exists in an ideal state represented by point A being on the steep portion of the curve). In cases in which there is a permanent loss or gain of the activity of a specific molecule (thin and thick curves, respectively), homeostatic changes in protein X occur to restore appropriate synaptic output (points B and C, respectively). At these new steady states, however, the neuron no longer retains the flexibility necessary to adequately respond to a learning or extinction stimulus (because points B and C are on flatter regions of the curve). Therefore, the dynamic ability of the neuronal system to respond is weakened, and phenotypes such as mental retardation and autism can result.

in the neurons. This is not surprising given the multitude of processes that contribute to homeostatic regulation of neuronal function and the large number of genes whose expression is altered on either loss or gain of the protein[53]. Within the context of our model we predict that changes induced by either loss or gain of protein function dynamically weaken the responsiveness of neuronal networks to stimuli (Fig. 3). This model could also explain the fluctuating nature of many psychiatric diseases.

Future challenges

To gain better insight into how neuron-specific defects compromise network function, improved techniques for studying neuronal networks *in vivo* are essential. It will also be crucial to evaluate whether and how the homeostatic capacity of a network is compromised following an initial compensatory process or in response to a primary defect. To determine whether multiple neuronal networks impinge on social behaviour, cognition and communication to cause a disorder such as autism, or whether subtle defects or changes within single networks are sufficient, will require specific investigations aimed at defining spatial and cell-specific threshold effects. Determining how to increase the synaptic flexibility of a network irrespective of the primary molecular defect might prove to be a fruitful way to manage the shared symptoms in this class of cognitive and behavioural disorders. Another therapeutic approach might involve dampening the homeostatic changes that occur in response to a primary defect.

The observation that functional loss or gain of proteins or RNAs involved in diverse processes leads to mental retardation, autism and other neuropsychiatric symptoms predicts that there are likely to be hundreds to thousands of genes whose alteration results in these common neurological phenotypes. A recent search by us of the Online Mendelian Inheritance in Man (OMIM) database yielded 305 genes known to be associated with mental retardation, 51 with autism and 75 with a behavioural disorder.

Another important prediction based on this theory of failed neuronal homeostasis in neuropsychiatric disease is that disorders such as schizophrenia, bipolar disorder, anxiety disorders and obsessive–compulsive disorder will have two main aetiologies. One will be genetic, with such disorders being caused by less severe alterations either in some of the genes or pathways that cause severe neurodevelopmental disorders such as autism (and whose effects often converge on the synapse) or in genes that have a crucial role in modulating synaptic plasticity in adolescence and into adulthood. The identification of individuals with significant psychiatric morbidity (for example, psychosis) among those with a mutation in *MECP2*, maternal uniparental disomy in PWS, or the 22q11.2 deletion syndrome supports this hypothesis. The second aetiology is likely to involve acquired epigenetic changes that result in altered gene expression or RNA processing, often in the same pathways that cause autism or similar neuropsychiatric disease in children. The fact that there are probably hundreds, if not thousands, of such genes may explain the poor yield of specific gene associations in psychiatric diseases, the small number of cases resulting from a defect in any one gene, and the variable penetrance and expressivity of these alleles in kindreds. It is probable that alteration of the same gene can cause anxiety in one individual, obsessive–compulsive disorder in another, psychosis in another, and autism in yet another. Undoubtedly, these disorders are modified by genetic background and life experience, through which synaptic plasticity and epigenetic changes may be either protective or predisposing to disease.

The identification of molecular pathways common to mental retardation and neuropsychiatric disorders will lead to better understanding of neuronal networks and the underpinnings of cognition, language, behaviour, social interaction and emotion. For therapeutic trials, it will be important to group patients by pathophysiological mechanism of disease rather than simply by which gene is affected. Finally, the identification of drugs that perturb or restore neuronal and synaptic homeostasis may significantly improve the quality of life for patients and their families. ■

1. Verkerk, A. J. *et al.* Identification of a gene (*FMR-1*) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65,** 905–914 (1991).
2. Garber, K., Smith, K. T., Reines, D. & Warren, S. T. Transcription, translation and fragile X syndrome. *Curr. Opin. Genet. Dev.* **16,** 270–275 (2006).
3. Comery, T. A. *et al.* Abnormal dendritic spines in fragile X knockout mice: maturation and pruning deficits. *Proc. Natl Acad. Sci. USA* **94,** 5401–5404 (1997).
4. Peier, A. M. *et al.* (Over)correction of *FMR1* deficiency with YAC transgenics: behavioral and physical features. *Hum. Mol. Genet.* **9,** 1145–1159 (2000).
5. Brown, V. *et al.* Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell* **107,** 477–487 (2001).
6. Darnell, J. C. *et al.* Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell* **107,** 489–499 (2001).
7. Bear, M. F., Huber, K. M. & Warren, S. T. The mGluR theory of fragile X mental retardation. *Trends Neurosci.* **27,** 370–377 (2004).
8. Dolen, G. *et al.* Correction of fragile X syndrome in mice. *Neuron* **56,** 955–962 (2007).
This paper provides genetic evidence that supports the mGluR theory of fragile X pathogenesis and proposes a potential therapeutic strategy.
9. Nakamoto, M. *et al.* Fragile X mental retardation protein deficiency leads to excessive mGluR5-dependent internalization of AMPA receptors. *Proc. Natl Acad. Sci. USA* **104,** 15537–15542 (2007).
10. Muddashetty, R. S., Kelic, S., Gross, C., Xu, M. & Bassell, G. J. Dysregulated metabotropic glutamate receptor-dependent translation of AMPA receptor and postsynaptic density-95 mRNAs at synapses in a mouse model of fragile X syndrome. *J. Neurosci.* **27,** 5338–5348 (2007).
This paper provides a mechanism underlying abnormal AMPA-receptor surface expression in excitatory synapses in fragile X syndrome and suggests that the key principle responsible for fragile X syndrome is that synaptic activation cannot stimulate the additional local protein synthesis necessary for synaptic plasticity to occur.
11. Yan, Q. J., Rammal, M., Tranfaglia, M. & Bauchwitz, R. P. Suppression of two major fragile X syndrome mouse model phenotypes by the mGluR5 antagonist MPEP. *Neuropharmacology* **49,** 1053–1066 (2005).
12. McBride, S. M. *et al.* Pharmacological rescue of synaptic plasticity, courtship behavior, and mushroom body defects in a *Drosophila* model of fragile X syndrome. *Neuron* **45,** 753–764 (2005).

13. Kobrynski, L. J. & Sullivan, K. E. Velocardiofacial syndrome, DiGeorge syndrome: the chromosome 22q11.2 deletion syndromes. *Lancet* **370,** 1443–1452 (2007).
14. Gothelf, D. *et al.* Risk factors for the emergence of psychotic disorders in adolescents with 22q11.2 deletion syndrome. *Am. J. Psychiatry* **164,** 663–669 (2007).
15. Lee, J. A. & Lupski, J. R. Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron* **52,** 103–121 (2006).
16. Ensenauer, R. E. *et al.* Microduplication 22q11.2, an emerging syndrome: clinical, cytogenetic, and molecular analysis of thirteen patients. *Am. J. Hum. Genet.* **73,** 1027–1040 (2003).
17. Yobb, T. M. *et al.* Microduplication and triplication of 22q11.2: a highly variable syndrome. *Am. J. Hum. Genet.* **76,** 865–876 (2005).
18. Paylor, R. *et al.* *Tbx1* haploinsufficiency is linked to behavioral disorders in mice and humans: implications for 22q11 deletion syndrome. *Proc. Natl Acad. Sci. USA* **103,** 7729–7734 (2006).
19. Long, J. M. *et al.* Behavior of mice with mutations in the conserved region deleted in velocardiofacial/DiGeorge syndrome. *Neurogenetics* **7,** 247–257 (2006).
20. Yagi, H. *et al.* Role of TBX1 in human del22q11.2 syndrome. *Lancet* **362,** 1366–1373 (2003).
21. Hiroi, N. *et al.* A 200-kb region of human chromosome 22q11.2 confers antipsychotic-responsive behavioral abnormalities in mice. *Proc. Natl Acad. Sci. USA* **102,** 19132–19137 (2005).
22. Paterlini, M. *et al.* Transcriptional and behavioral interaction between 22q11.2 orthologs modulates schizophrenia-related phenotypes in mice. *Nature Neurosci.* **8,** 1586–1594 (2005).
23. Stark, K. L. *et al.* Altered brain microRNA biogenesis contributes to phenotypic deficits in a 22q11-deletion mouse model. *Nature Genet.* **40,** 751–760 (2008).
This paper suggests a novel pathophysiological mechanism for the cognitive and psychiatric phenotypes observed in the human 22q11.2 deletion syndrome: abnormal miRNA biogenesis.
24. Lalande, M. & Calciano, M. A. Molecular epigenetics of Angelman syndrome. *Cell Mol. Life Sci.* **64,** 947–960 (2007).
25. Battaglia A. The inv dup(15) or idic(15) syndrome: a clinically recognisable neurogenetic disorder. *Brain Dev.* **27,** 365–369 (2005).
26. Sahoo, T. *et al.* Prader–Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nature Genet.* **40,** 719–721 (2008).
This paper confirms the cause of PWS — deficiency of non-coding RNA molecules important for normal RNA processing — and suggests a novel role for snoRNAs in cognitive and psychiatric disease.
27. Miura, K. *et al.* Neurobehavioral and electroencephalographic abnormalities in *Ube3a* maternal-deficient mice. *Neurobiol. Dis.* **9,** 149–159 (2002).
28. Gallagher, R. C., Pils, B., Albalwi, M. & Francke, U. Evidence for the role of PWCR1/HBII-85 C/D box small nucleolar RNAs in Prader–Willi syndrome. *Am. J. Hum. Genet.* **71,** 669–678 (2002).
29. Ding, F. *et al.* SnoRNA *Snord116* (*Pwcr1/MBII-85*) deletion causes growth deficiency and hyperphagia in mice. *PLoS ONE* **3,** e1709 (2008).
30. Elsea, S. H. & Girirajan, S. Smith–Magenis syndrome. *Eur. J. Hum. Genet.* **16,** 412–421 (2008).
31. Girirajan, S. *et al.* How much is too much? Phenotypic consequences of *Rai1* overexpression in mice. *Eur. J. Hum. Genet.* **16,** 941–954 (2008).
32. Smith, A. C. *et al.* Interstitial deletion of (17)(p11.2p11.2) in nine patients. *Am. J. Med. Genet.* **24,** 393–414 (1986).
33. Chen, K. S. *et al.* Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome. *Nature Genet.* **17,** 154–163 (1997).
34. Slager, R. E., Newton, T. L., Vlangos, C. N., Finucane, B. & Elsea, S. H. Mutations in *RAI1* associated with Smith–Magenis syndrome. *Nature Genet.* **33,** 466–468 (2003).
35. Imai, Y. *et al.* Cloning of a retinoic acid-induced gene, *GT1*, in the embryonal carcinoma cell line P19: neuron-specific expression in the mouse brain. *Brain Res. Mol. Brain Res.* **31,** 1–9 (1995).
36. Bi, W. *et al.* Mutations of *RAI1*, a PHD-containing protein, in nondeletion patients with Smith–Magenis syndrome. *Hum. Genet.* **115,** 515–524 (2004).
37. Bi, W. *et al.* Inactivation of *Rai1* in mice recapitulates phenotypes observed in chromosome engineered mouse models for Smith–Magenis syndrome. *Hum. Mol. Genet.* **14,** 983–995 (2005).
38. Bi, W. *et al.* *Rai1* deficiency in mice causes learning impairment and motor dysfunction, whereas *Rai1* heterozygous mice display minimal behavioral phenotypes. *Hum. Mol. Genet.* **16,** 1802–1813 (2007).
39. Potocki, L. *et al.* Molecular mechanism for duplication 17p11.2 — the homologous recombination reciprocal of the Smith–Magenis microdeletion. *Nature Genet.* **24,** 84–87 (2000).
40. Potocki, L. *et al.* Characterization of Potocki–Lupski syndrome (dup(17)(p11.2p11.2)) and delineation of a dosage-sensitive critical interval that can convey an autism phenotype. *Am. J. Hum. Genet.* **80,** 633–649 (2007).
41. Walz, K., Paylor, R., Yan, J., Bi, W. & Lupski, J. R. *Rai1* duplication causes physical and behavioral phenotypes in a mouse model of dup(17)(p11.2p11.2). *J. Clin. Invest.* **116,** 3035–3041 (2006).
42. Moretti, P. & Zoghbi, H. Y. MeCP2 dysfunction in Rett syndrome and related disorders. *Curr. Opin. Genet. Dev.* **16,** 276–281 (2006).
43. Chahrour, M. & Zoghbi, H. Y. The story of Rett syndrome: from clinic to neurobiology. *Neuron* **56,** 422–437 (2007).
44. Amir, R. E. *et al.* Rett syndrome is caused by mutations in X-linked *MECP2*, encoding methyl-CpG-binding protein 2. *Nature Genet.* **23,** 185–188 (1999).
45. del Gaudio, D. *et al.* Increased *MECP2* gene copy number as the result of genomic duplication in neurodevelopmentally delayed males. *Genet. Med.* **8,** 784–792 (2006).
46. Van Esch, H. *et al.* Duplication of the *MECP2* region is a frequent cause of severe mental retardation and progressive neurological symptoms in males. *Am. J. Hum. Genet.* **77,** 442–453 (2005).
47. Meins, M. *et al.* Submicroscopic duplication in Xq28 causes increased expression of the *MECP2* gene in a boy with severe mental retardation and features of Rett syndrome. *J. Med. Genet.* **42,** e12 (2005).
48. Friez, M. J. *et al.* Recurrent infections, hypotonia, and mental retardation caused by duplication of *MECP2* and adjacent region in Xq28. *Pediatrics* **118,** e1687–e1695 (2006).
49. Smyk, M. *et al.* Different-sized duplications of Xq28, including *MECP2*, in three males with mental retardation, absent or delayed speech, and recurrent infections. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **147B,** 799–806 (2008).
50. Guy, J., Hendrich, B., Holmes, M., Martin, J. E. & Bird, A. A mouse *Mecp2*-null mutation causes neurological symptoms that mimic Rett syndrome. *Nature Genet.* **27,** 322–326 (2001).
51. Collins, A. L. *et al.* Mild overexpression of *MeCP2* causes a progressive neurological disorder in mice. *Hum. Mol. Genet.* **13,** 2679–2689 (2004).
52. Samaco, R. C. *et al.* A partial loss of function allele of methyl-CpG-binding protein predicts a human neurodevelopmental syndrome. *Hum. Mol. Genet.* **17,** 1718–1727 (2008).
53. Chahrour, M. *et al.* MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* **320,** 1224–1229 (2008).
54. Yasui, D. H. *et al.* Integrated epigenomic analyses of neuronal MeCP2 reveal a role for long-range interaction with active genes. *Proc. Natl Acad. Sci. USA* **104,** 19416–19421 (2007).
55. Chao, H.-T., Zoghbi, H. & Rosenmund, C. MeCP2 controls excitatory synaptic strength by regulating glutamatergic synapse number. *Neuron* **56,** 58–65 (2007).
This paper provides evidence that either loss or gain of MeCP2 alters excitatory synaptic function, leading to overlapping abnormal neurological phenotypes.
56. Zhou, Z. *et al.* Brain-specific phosphorylation of MeCP2 regulates activity-dependent *Bdnf* transcription, dendritic growth, and spine maturation. *Neuron* **52,** 255–269 (2006).
57. Martinowich, K. *et al.* DNA methylation-related chromatin remodeling in activity-dependent *BDNF* gene regulation. *Science* **302,** 890–893 (2003).
58. McGill, B. E. *et al.* Enhanced anxiety and stress-induced corticosterone release are associated with increased *Crh* expression in a mouse model of Rett syndrome. *Proc. Natl Acad. Sci. USA* **103,** 18267–18272 (2006).
59. Bourgeron, T. The possible interplay of synaptic and clock genes in autism spectrum disorders. *Cold Spring Harb. Symp. Quant. Biol.* **72,** 645–654 (2007).
60. Berg, J. S. *et al.* Speech delay and autism spectrum behaviors are frequently associated with duplication of the 7q11.23 Williams–Beuren syndrome region. *Genet. Med.* **9,** 427–441 (2007).
61. Somerville, M. J. *et al.* Severe expressive-language delay related to duplication of the Williams–Beuren locus. *N. Engl. J. Med.* **353,** 1694–1701 (2005).
62. Torniero, C. *et al.* Cortical dysplasia of the left temporal lobe might explain severe expressive-language delay in patients with duplication of the Williams–Beuren locus. *Eur. J. Hum. Genet.* **15,** 62–67 (2007).
63. Ewart, A. K. *et al.* Hemizygosity at the elastin locus in a developmental disorder, Williams syndrome. *Nature Genet.* **5,** 11–16 (1993).
64. Tassabehji, M. Williams–Beuren syndrome: a challenge for genotype–phenotype correlations. *Hum. Mol. Genet.* **12** (special no. 2), R229–R237 (2003).
65. Zhao, C. *et al.* Hippocampal and visuospatial learning defects in mice with a deletion of frizzled 9, a gene in the Williams syndrome deletion interval. *Development* **132,** 2917–2927 (2005).
66. Hoogenraad, C. C. *et al.* Targeted mutation of *Cyln2* in the Williams syndrome critical region links *CLIP-115* haploinsufficiency to neurodevelopmental abnormalities in mice. *Nature Genet.* **32,** 116–127 (2002).
67. Meng, Y. *et al.* Abnormal spine morphology and enhanced LTP in LIMK-1 knockout mice. *Neuron* **35,** 121–133 (2002).
68. Heredia, L. *et al.* Phosphorylation of actin-depolymerizing factor/cofilin by LIM-kinase mediates amyloid β-induced degeneration: a potential mechanism of neuronal dystrophy in Alzheimer's disease. *J. Neurosci.* **26,** 6533–6542 (2006).
69. Lim, M. K. *et al.* Parkin interacts with LIM kinase 1 and reduces its cofilin-phosphorylation activity via ubiquitination. *Exp. Cell Res.* **313,** 2858–2874 (2007).

Acknowledgements In memory of our mentor, Ralph D. Feigin. We are grateful to C. Rosenmund for careful reading of the manuscript, discussions and helping us to articulate our hypothesis. We are indebted to the Howard Hughes Medical Institute, the National Institute of Neurological Disorders and Stroke (grant number 1R01 NS057819-01 to H.Y.Z., and grant numbers T32 NS43124 and 1K08 NS062711-01 to M.B.R.) and the Simons Foundation for supporting our research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests.
Correspondence should be addressed to the authors (mramocki@bcm.tmc.edu; hzoghbi@bcm.tmc.edu).

Copy-number variations associated with neuropsychiatric conditions

Edwin H. Cook Jr[1] & Stephen W. Scherer[2,3]

Neuropsychiatric conditions such as autism and schizophrenia have long been attributed to genetic alterations, but identifying the genes responsible has proved challenging. Microarray experiments have now revealed abundant copy-number variation — a type of variation in which stretches of DNA are duplicated, deleted and sometimes rearranged — in the human population. Genes affected by copy-number variation are good candidates for research into disease susceptibility. The complexity of neuropsychiatric genetics, however, dictates that assessment of the biomedical relevance of copy-number variants and the genes that they affect needs to be considered in an integrated context.

Rapid developments in microarray technologies and the discovery of extensive copy-number variation (CNV) in the human genome[1,2] have provided new opportunities to examine the range of genetic variants associated with human disease. CNVs are segments of DNA ranging from 1 kilobase (kb) to several megabases (Mb), for which copy-number differences have been revealed by comparison of two or more genomes[3]. These quantitative structural variants often encompass one or more genes and collectively include more nucleotides per genome than the total number of single-nucleotide polymorphisms (SNPs)[4]. They can be copy-number gains (duplications or insertional transpositions), losses (deletions), gains or losses of the same locus, or multiallelic or complex rearrangements[4].

CNVs can affect gene function in several ways. Deletion or disruption of one or more genes can cause functional loss, with CNVs functioning as dominant or recessive alleles according to the cellular function of the affected gene product(s). Alternatively, CNVs may disrupt a regulatory element, generate novel fusion products, or act by way of position effect, with various possible positive and negative sequelae, including imprinting and differential allelic expression[3,5]. Conversely, some large CNVs involving dozens of genes have no overt phenotype[1–4].

We are just beginning to recognize CNV features that are relevant to medical research. First, family data suggest a comparatively high locus-specific rate of new nucleotide mutation in CNVs, three to four times that for SNPs[6,7]. Second, some balanced CNVs with no functional effect in a carrier can create genomic instability in future generations[3]. Third, larger CNVs (>50 kb) are distributed non-uniformly along chromosomes and are relatively rare, suggesting that they may be shaped by purifying selection (with deletions under stronger selective pressure than duplications)[4,8]. Last, syndromes or phenotypes result from the alteration of specific genomic segments, such as the many dosage-related microdeletions and duplications that cause genomic disorders[9]. CNV 'mutations' have been found among all types of Mendelian trait and a growing number of complex disorders[10], including those that manifest as maladies of the nervous system[11].

In this Review, we focus on recent genome-wide studies testing for associations of CNV with neuropsychiatric disease, and on the broader impact of CNVs. Data from these studies are providing a glimpse into the genetic architecture underlying neuropsychiatric disease and revealing many new candidate disease-susceptibility genes.

CNV genomics in neuropsychiatric disease

Table 1 summarizes examples of conditions with neuropsychiatric features for which germline CNVs are involved. Other reviews have summarized the role of chromosomal abnormalities and candidate genes in the broader groupings of neurodevelopment, and neurodegenerative and psychiatric disorders[11–13].

The mechanisms by which CNV can cause neuropsychiatric disorders are perhaps most clearly shown by those disorders caused by rare and typically (but not always) highly penetrant CNVs, involving the loss, gain or disruption of a dosage-sensitive gene(s)[11,12]. The resultant phenotypes are consistent enough to be grouped into syndromes, but there can be variable expression of features within each grouping. For example, among patients with the 22q11.2 deletion syndrome, which is associated with a 3-Mb microdeletion, approximately 25% have psychiatric manifestations that include schizophrenia, attention-deficit hyperactivity disorder (ADHD) or autism spectrum disorders (ASDs). Williams–Beuren syndrome (WBS) — caused by hemizygosity of around 20 genes at 7q11.23 — is associated with visuospatial deficits, a spectrum of learning disabilities, and ADHD in 70% of cases. Moreover, the type of CNV rearrangement and whether it is maternally or paternally inherited can influence the phenotype. For example, maternally derived duplications of chromosome 15q11–q13 confer a high risk of ASD (>85%), in contrast to paternally derived duplications whose effects can range from no effect to mild developmental and cognitive impairment[14,15].

In most genomic disorders, such as those involving 22q11.2, the WBS region or 15q11–q13, the CNV interval is flanked by highly homologous low-copy-number repeats (LCRs), which can be substrates for recombination between non-allelic LCRs (non-allelic homologous recombination), resulting in deletion or duplication of the same intervening segment[9]. The clinical presentation associated with deletions tends to be more uniform and commonly includes dysmorphology, which facilitates diagnosis. For duplications, however, evidence is emerging for a more subtle presentation in many cases, creating the potential for duplication carriers to go undetected. As a result, some of these genomic disorders,

[1]Institute for Juvenile Research, Department of Psychiatry, University of Illinois, 1747 West Roosevelt Road, Chicago, Illinois 60608, USA. [2]The Centre for Applied Genomics, Hospital for Sick Children, 14th floor, Toronto Medical Discovery Tower/MaRS Discovery District, 101 College Street, Toronto, Ontario M5G 1L7, Canada. [3]Department of Molecular Genetics, University of Toronto, Room 4396, 1 King's Circle, Toronto, Ontario M5S 1A8, Canada.

such as the WBS duplications, are under-diagnosed[16,17]. The corollary, however, is that CNVs involving duplication are under less selective pressure, and thus may be maintained longer in the population.

Most of the severe psychiatric disorders are highly heritable (for example, ASD, bipolar disorder and schizophrenia are >80% heritable)[12,13]. A trend in the investigation of genetic influences has been to test the 'common disease–common variant' model, which proposes that multiple common alleles in the population contribute small-to-moderate additive or multiplicative effects to the phenotype. Such approaches have been facilitated by advanced methods for genotyping SNPs, but although they have dominated the field, they have yielded only a handful of candidate genes[13]. Because of technical limitations, there have yet to be proper genome-wide association studies testing the role of common CNVs in neuropsychiatric disease. Conversely, the 'common disease–rare variant' hypothesis posits that many alternative rare variants underlie susceptibility to common complex conditions. Such SNPs or CNVs would be rare (perhaps unique to a family), of recent origin and often highly penetrant.

CNVs in autism spectrum disorder and schizophrenia

Evidence is accumulating that multiple rare *de novo* (and some inherited) CNVs contribute to the genetic component of vulnerability to ASD or schizophrenia. Some karyotypic and submicroscopic CNV abnormalities have been known for some time, including the maternal 15q11–q13 duplication found in 1–3% of individuals with ASD[18]. In some syndromes — such as Potocki–Lupski syndrome, which is caused by a duplication of 17p11.2 — more than 90% of cases show features of ASD[15]. The CNVs observed in individuals with ASD or ASD-related

Table 1 | Examples of conditions with neuropsychiatric features associated with CNVs

Condition	OMIM entry no.	Gene(s)†	Locus	CNV features‡
Genomic disorders (microdeletion or duplication)				
22q11.2 deletion syndrome (DiGeorge syndrome and velocardiofacial syndrome)	#188400, #192430	*TBX1* +	22q11.2	Contiguous gene deletion
22q11.2 duplication syndrome	#608363	+++	22q11.2	Diverse phenotypes, even within families (including in unaffected carriers)
Miller–Dieker lissencephaly syndrome	#247200	*LIS1* +	17p13.3	Contiguous gene deletion
Neurofibromatosis type 1	+162200	*NF1* +	17q11.2	5–20% Microdeletions involving contiguous genes (syndromic form)[45]
	+162200	*NF1* +	17q11.2	Microduplication (one family)[46]
Pelizaeus–Merzbacher disease	#312080	*PLP1*	Xq22	X-linked recessive, non-recurrent duplication or deletion[47]
Prader–Willi syndrome	#176270	+++	15q11–q13	Paternal contiguous gene deletion
Angelman syndrome	#105830	*UBE3A* +	15q11–q13	Maternal contiguous gene deletion
15q11–q13 duplication syndrome (see also Autism)§	%608636	+++	15q11–q13	Maternal contiguous gene duplication
Smith–Magenis syndrome	#182290	*RAI1* +	17p11.2	Contiguous gene deletion
Potocki–Lupski syndrome	#610883	+++	17p11.2	Non-recurrent contiguous gene duplication
Williams–Beuren syndrome (WBS)	#194050	+++	7q11.23	Recurrent contiguous gene deletion
Williams–Beuren-region duplication syndrome	#609757	+++	7q11.23	Recurrent contiguous gene duplication; CNV also in controls
Monogenic and complex disease				
Alzheimer's disease (early onset)	#104300, *104760	*APP*	21q21	Dominant gene-dosage effect
Autism	#606232	*SHANK3*	22q13	Rare deletions
Autism spectrum disorder (ASD)	#300495	*NLGN4*	Xp22.33	Rare deletions
	%611913	+++	16p11.2	Contiguous gene deletion or duplication (~1% of cases)
	*608286	*PCDH10*¶	4q28.3	Rare homozygous deletion[26]
	*608396	*NHE9*¶	3q24	Rare homozygous deletion[26]
	%608636	+++	15q11–q13	Maternal duplication (1–3% of cases)
	–	+++	Multiple	Multiple rare *de novo* CNVs (Fig. 1)
Bipolar disorder	*605004	*GSK3B*	3q13.3	CNV gains exceed losses of candidate gene (one study)[48]
Rett syndrome	#312750	*MECP2*	Xq28	X-linked recessive, rare intragenic deletions (duplications associated with atypical presentation)
Parkinson's disease	#168601	*SNCA*	4q22	Multi-allelic dosage effects[49,50]
Schizophrenia	#181500	*ERBB4*¶	2q34	Rare deletion[27]
	#181500	*SLC1A3*¶	5p13.3	Rare deletion[27]
	#181500	*RAPGEF4*¶	2q31.2	Rare deletion[28]
	#181500	*CIT*¶	12q24	Rare duplication[28]
	–	+++	Multiple	Multiple rare *de novo* CNVs[27–30]
	–	+++	1q21.1, 15q11.2 and 15q13.3	Rare recurrent CNVs[29,30]
Tourette's syndrome	#137580	Unspecified	17p11.2	Microdeletion (one case)[51]
	#137580	*NLGN4*	Xp22.33	Intragenic deletion (one family)[52]

†A gene symbol followed by a + indicates that this gene is the predominant phenotype-determining gene in a multi-gene CNV. +++ indicates that CNVs contain multiple genes. ‡References cited are those not listed in the Online Mendelian Inheritance in Man (OMIM) database. §Note that the 15q11–q13 duplication syndrome appears as entries in both the Genomic disorder row and the Autism row, exemplifying that the classification of a CNV can depend on whether it is discovered in a genotype-driven study or a phenotype-driven study[10] (Fig. 2). ¶These genes were identified by genome-wide screens and therefore should be considered only as candidates for disease susceptibility, as they require further study.

syndromes — an area of our research from which we draw examples — are summarized in Fig. 1.

To quantify the role of CNV in ASD, a series of studies used different microarray platforms based on oligonucleotides[19], SNPs[20–22] or clones[23–25] to examine cohorts of idiopathic cases (that is, cases with no known cause). The resolution of detection was better than with karyotyping but was still relatively crude (median CNV ~500 kb). The families examined had either one child or more (simplex and multiplex families, respectively) who met minimal standard criteria for an ASD (autism spectrum disorder, pervasive developmental disorder or Asperger's syndrome).

Relevant findings included the following. First, the proportion of *de novo* CNVs differed between simplex families (7–10%), multiplex families (2–3%) and non-ASD controls (1%)[19,21], similarly to other complex inherited disorders that show evidence for different mechanisms underlying sporadic and familial cases. Second, some individuals with ASD had two or more *de novo* CNVs[21,24], and 27% of families with syndromic forms of ASD carried *de novo* events[25]. Third, up to 40% of family-specific CNVs were found to be inherited from a parent who did not seem to have ASD[20,21,24]. Fourth, in several examples, unrelated individuals carried *de novo* or inherited CNVs at the same locus, and some of these coincided with known genomic disorders[19–25] (Fig. 1 and Table 1), whereas others involved genes associated with mental retardation[21,25]. In three studies, the most common region identified (apart from maternal 15q11–q13 duplication) was an approximately 600-kb microdeletion or microduplication of chromosome 16p11.2 (in ~1% of cases)[21–23], and these CNVs were also observed in ASD cases with additional dysmorphology[21–23], in non-ASD cases having developmental delay or mental retardation[22], and in controls[21–23]. Last, there was a relative enrichment within CNVs for neuronal synaptic complex genes, notably *SHANK3*, *NLGN4* and *NRXN1* (refs 19–24). Another study of consanguineous ASD families identified rare homozygous CNVs deleting both copies of genes such as *PCDH10*, *DIA1* and *NHE9* (ref. 26), all genes regulated by neuronal activity.

Genome-wide CNV data concerning schizophrenia are also available. One case–control study[27] found more novel and rare CNVs in schizophrenia cases with adult (15%) or childhood (20%) onset than in controls (5%). The rate of new mutation was not determined. Another study[28] tested for association of *de novo* CNVs with schizophrenia, finding a frequency of 10% in sporadic cases, 1.3% in unaffected controls and none in familial cases (noting that 'none' may reflect the small sample size). Similar to findings for ASD[21,24], around 1% of schizophrenia cases carried two or more *de novo* CNV events[28]. Notable candidate genes for schizophrenia included *ERBB4* (ref. 27), *SLC1A3* (ref. 27), *RAPGEF4* (ref. 28) and *CIT*[28] (which, from its location, is also a candidate for bipolar disorder), as well as *NRXN1* (ref. 27) and the 16p11.2 region[27] discussed earlier (the latter two loci being found in studies of ASD and mental retardation[20–24]).

Two other studies examined CNVs in ancestrally matched schizophrenia cases and controls[29] or large recurrent CNVs proposed to be under negative selection because of reduced fecundity associated with schizophrenia[30]. Hemizygous *de novo* deletions at chromosomes 1q21.1 (>1 Mb multi-allelic CNVs)[29,30], 15q11.2 (470 kb)[30] and 15q13.3 (~1.5 Mb)[29,30], all involving flanking LCRs and numerous genes, were most often identified, albeit at low frequencies (<1%). These CNV loci had also previously been observed in other patients (with no diagnosis of schizophrenia). The locus 1q21.1, for example, is associated with duplications in autism[20], deletions or duplications in mental retardation[31–33], and deletions or duplications in congenital cardiac defects and lens abnormalities[4,33]. The 15q13.3 locus is associated with deletions in mental retardation with seizures[34], and deletions or duplications in autism and other neuropsychiatric disorders[35].

Determining causation

Neuropsychiatric genetics is awash with CNV studies — a reason for great optimism, but also for caution in extrapolating from research results, particularly to clinical applications[10,36]. In Fig. 2, we present a series of questions that need to be considered when attempting to assess the relevance of a CNV(s) in research and medical practice.

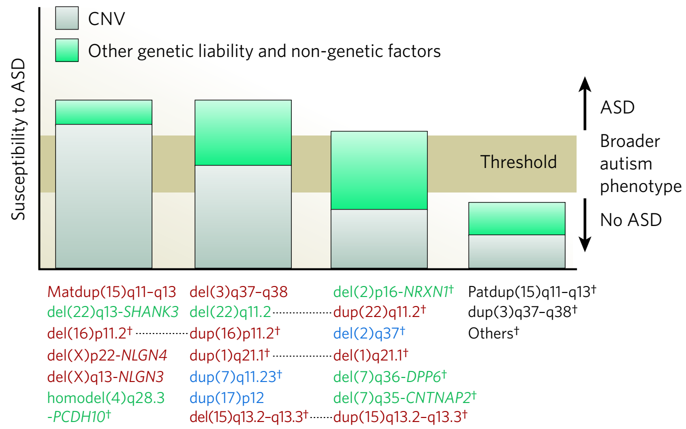


Figure 1 | Model of the relative contribution of CNVs in susceptibility to ASD. CNVs (and the accompanying dosage-sensitive genes) associated with ASD pathogenesis are ordered from left to right on the basis of observed relative penetrance. In the columns, grey is the contribution of CNVs; green is the contribution of other genetic liability and non-genetic factors. The broader autism phenotype (threshold) bar represents the graded expression of genetic factors (some predisposing, some possibly protective) that influence liability to the principal attributes of ASD (that is, qualitative impairments in social-communication function and behavioural flexibility). Such factors are expressed most strongly among parents from multiplex families and less strongly among simplex families. The 4:1 male-to-female ratio in ASD suggests that penetrance is lower in females than in males. The first column represents CNVs with a potentially large impact on ASD susceptibility. These germline mutations would typically be *de novo* in origin, cause more severe ASD symptoms, be more prevalent among sporadic forms of ASD, and be less influenced by other factors such as gender; all of these trends have been observed[19–25]. The second and third columns represent CNVs with moderate or mild effects that probably require other genetic (or non-genetic) factors to take the phenotype across the ASD threshold. The relative contribution of CNVs to ASD susceptibility is an estimate based on the literature and will vary between individuals for the reasons discussed here and in Fig. 2. CNVs denoted with a dagger (†) are found in non-ASD family members (including hemizygous CNV carriers for *PCDH10*)[26], other disorders without ASD, or population controls, consistent with this model. A gene name is included with the CNV (for example, del(22)q13-*SHANK3*) if there is supporting evidence that the gene is involved (such as CNV recurrence in unrelated families or mutation detection in individuals without the CNV). The primary referral diagnosis for ASD can be idiopathic (for CNVs listed in green), an ASD-related syndrome (blue; for example, duplication of 17p12-*RAI1* in Potocki–Lupski syndrome or duplication of 7q11.23), or ASD or ASD-related in different individuals with the same CNV genotype (red). Black denotes no diagnosis of ASD. Dotted lines identify reciprocal CNV gains or losses of the same locus. Some CNV loci (1q21.1, 15q13.2–q13.3, 16p11.2 and 22q11.2) are associated with variable phenotypes. del, deletion; dup, duplication; Mat, maternal; Pat, paternal.

At present, it is difficult to know whether a given *de novo* variant is pathological, and part of the reason for this is our rudimentary knowledge of mutation rates and population-distribution statistics of CNVs. It is also difficult to know whether an inherited variant is necessarily benign in a particular genomic environment. For example, the extent of the CNV and its inclusion of entire genes (or of specific isoforms of genes) may influence the phenotype, as can nearby genes[37]. Moreover, some structural variants annotated as benign may, on retrospective evaluation, be found to be associated with subtle phenotypic signs[21,23]. All these issues are exacerbated in studies of complex disease, where the genetic liability of other risk alleles or protective modifiers can influence expressivity and penetrance[10] (Figs 1 and 2).

De novo CNVs are found in the general population at comparatively high frequencies[6,7], sometimes two or more per sample. As just one example[21], an individual with ASD carried a 277-kb deletion of *SHANK3*, as well as a larger 1.4-Mb duplication (including 44 genes) on chromosome 20q13. *SHANK3* is known, from other rigorous studies involving mutation screening, to be a rare monogenic cause of ASD[38]; therefore, it

is most probable that the *SHANK3 de novo* deletion caused ASD in this case. Had the *SHANK3* alteration gone undetected, the larger *de novo* change might have been inappropriately implicated in ASD (although it may still influence the phenotype).

CNV genotype-to-phenotype studies have other complexities. Gains and losses involving several genes at the same locus (such as the loci 16p11.2 and 1q21.1) can lead to overlapping phenotypes or different phenotypes, sometimes with variable penetrance. In multiplex families with ASD, CNV (for example, at the 16p11.2 and 22q11.2 loci) has been seen in some affected siblings and not in others[21,23], suggesting threshold models of liability (Fig. 1). This is not surprising, as a '*de novo*' CNV should not recur in an affected sibling in the absence of post-zygotic mosaicism, an example being a CNV in *NRXN1* in a sibling pair with ASD[20]. Adding to the complexity, a recent comparison of CNVs in monozygotic twins[39] draws attention to the prevalence of somatic events that can create mosaicism in different tissues (such as brain), with possible contribution to phenotype.

Other independent potential risk factors, such as those that are genetic (including CNVs predisposing individuals to subsequent chromosomal abnormalities), epigenetic, sex-related, environmental or stochastic in origin may also warrant attention. All of these phenomena need to be considered in genetic counselling, and this will be particularly important for complex disease, for which the opportunity to use CNV data for diagnosis will often precede the ability to predict specific outcomes.

Outlook

Knowledge of CNV can assist researchers to stratify cohorts in genetic linkage and association studies[20], to build new models for testing genetic hypotheses[40,41], to identify candidate susceptibility genes, and to provide valuable data in the clinical genetics setting[42], in particular when new or known syndromes are being identified. Methods are now sufficiently sensitive (at least for detecting larger variants) and inexpensive that it would seem prudent to complement all genome-wide research studies of neuropsychiatric disease with a CNV-analysis step.

Most of the current data, and therefore our discussion here, concerns rare *de novo* variants with potentially large effects on phenotype (Fig. 1). Familial CNVs should also be flagged for consideration as possible aetiological factors in neuropsychiatric disease, in particular if they are rare in the general population[10,30]. Perhaps the most intriguing CNV observation is that common genes or pathways span different neuropsychiatric disorders (reminiscent of common SNP-association loci being found in different inflammatory and immune disorders[43]). Overlap of CNVs found in schizophrenia and ASD may be due to chance — approximately 1% of children with ASD develop schizophrenia, which is not a higher rate than in the general population. Alternatively, it could be due to overlap in symptoms across psychiatric diagnostic categories, misdiagnosis of ASD as schizophrenia, or pleiotropy and/or variable expressivity of the genes involved.

So far, data indicate that rare CNVs account for a proportion of the genetic liability to common complex and severe psychiatric disorders such as ASD and schizophrenia. Further systematic research is anticipated concerning the contribution of rare CNVs to bipolar disorder, obsessive–compulsive disorder, Tourette's syndrome, ADHD and other neuropsychiatric conditions. Because of technical issues of experimental noise and inability to resolve relative copy numbers, in particular at multi-allelic CNV loci[44], it has been difficult to carry out robust genome-wide association studies of CNVs[22,43]. New types of microarray are becoming available that should support reliable screening for common CNVs and higher-resolution genome coverage, allowing detection of even smaller CNVs. With these new tools, and with the rapid advancement in whole-genome sequencing technologies, a better understanding of the relative contribution of common and rare CNVs and SNPs in complex neuropsychiatric disease should surface. A challenge, particularly for complex traits, will be to determine how

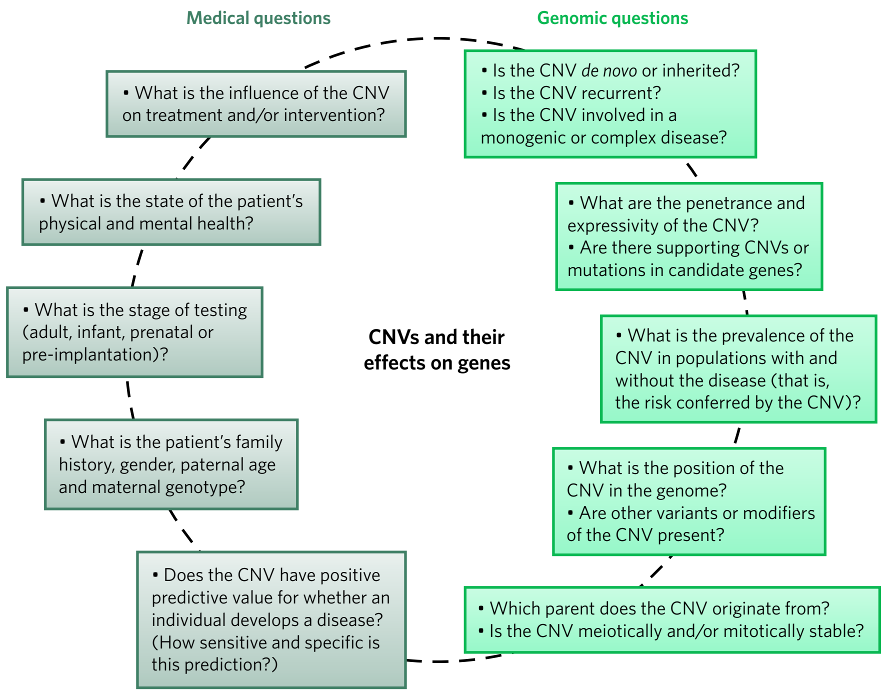


Figure 2 | Assessing the relevance of CNV. The characteristics of the CNV(s) in an individual's genome, including their effects on gene dosage, will affect the risk that the individual will develop particular diseases and the way that these diseases manifest. Questions that typically arise from genomic investigations (green) and medical investigations (grey) are shown. When considering the utility of a CNV as a diagnostic tool, the ideal strategy is to answer as many of these questions as possible and then to integrate the results. How these results are applied to research or to medical decision-making can be weighted according to the circumstances. For example, the relevance of the results will differ if a CNV is uncovered in the following four cases: one, a CNV is found in a gene known to cause disease or a CNV with a specific parent-of-origin or gender effect is found; two, a CNV is found in a high-risk setting (such as during prenatal complications or in a gene that might result in an adverse drug reaction); three, a CNV is found through a targeted test (such as in an individual with a family history of a particular disease or as confirmation of an existing clinical diagnosis); and, four, a CNV is found in a universal population screen.

CNVs and other variants interact, and to develop new models to extract these data.

Of the various types of structural genomic variation, we have discussed only CNV, as it is the most prevalent and is readily detected by microarrays. However, balanced alterations such as translocations and inversions can also be pathogenic[3], and for these, karyotyping is the assay of choice at present. Microarrays now detect all of the unbalanced chromosomal changes traditionally observed under the microscope[42]. When both of these methods were applied to the same ASD samples[21], it was found that of the ~7.4% of cases that carried *de novo* structural variants, only 0.4% carried balanced translocations or inversions. This suggests that in research and clinical settings, microarrays should become the first choice of screening technique — at least for ASD.

CNVs will continue to be discovered in the genomes of individuals with neuropsychiatric syndromes and threshold traits such as ASDs. As CNV becomes better understood and is coupled with more refined clinical descriptions, researchers and clinicians will find more susceptibility genes with individually small incremental effects. CNVs are also likely to be increasingly associated with characteristics that are continuously variable, change over time, or are cross-sectional in different clinical disorders. The comparison of such corresponding 'dimensional' phenotypic and genomic variation data promises to add new strata to our understanding of the influence of genes in human behaviour and disease. ■

1. Iafrate, A. J. *et al.* Detection of large-scale variation in the human genome. *Nature Genet.* **36,** 949–951 (2004).
2. Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305,** 525–528 (2004).
3. Feuk, L., Carson, A. R. & Scherer, S. W. Structural variation in the human genome. *Nature Rev. Genet.* **7,** 85–97 (2006).
4. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444,** 444–454 (2006).
This paper provides the first comprehensive CNV map of the human genome, describing the ubiquity and role of this form of variation in genetic diversity.
5. Lupski, J. R. & Stankiewicz, P. Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS Genet.* **1,** e49 (2005).
6. van Ommen, G. J. Frequency of new copy number variation in humans. *Nature Genet.* **37,** 333–334 (2005).
7. Lupski, J. R. Genomic rearrangements and sporadic disease. *Nature Genet.* **39,** S43–S47 (2007).
8. Emerson, J. J., Cardoso-Moreira, M., Borevitz, J. O. & Long, M. Natural selection shapes genome-wide patterns of copy-number polymorphism in *Drosophila melanogaster*. *Science* **320,** 1629–1631 (2008).
9. Stankiewicz, P. & Lupski, J. R. Genome architecture, rearrangements and genomic disorders. *Trends Genet.* **18,** 74–82 (2002).
10. Buchanan, J. A. & Scherer, S. W. Contemplating effects of genomic structural variation. *Genet. Med.* **10,** 639–647 (2008).
11. Lee, J. A. & Lupski, J. R. Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron* **52,** 103–121 (2006).
12. Inoue, K. & Lupski, J. R. Genetics and genomics of behavioral and psychiatric disorders. *Curr. Opin. Genet. Dev.* **13,** 303–309 (2003).
13. Burmeister, M., McInnis, M. G. & Zollner, S. Psychiatric genetics: progress amid controversy. *Naure Rev. Genet.* **9,** 527–540 (2008).
14. Veenstra-Vanderweele, J., Christian, S. L. & Cook, E. H. Jr. Autism as a paradigmatic complex genetic disorder. *Annu. Rev. Genomics Hum. Genet.* **5,** 379–405 (2004).
15. Abrahams, B. S. & Geschwind, D. H. Advances in autism genetics: on the threshold of a new neurobiology. *Nature Rev. Genet.* **9,** 341–355 (2008).
16. Turner, D. J. *et al.* Germline rates of *de novo* meiotic deletions and duplications causing several genomic disorders. *Nature Genet.* **40,** 90–95 (2008).
17. Berg, J. S. *et al.* Speech delay and autism spectrum behaviors are frequently associated with duplication of the 7q11.23 Williams–Beuren syndrome region. *Genet. Med.* **9,** 427–441 (2007).
18. Cook, E. H. Jr *et al.* Autism or atypical autism in maternally but not paternally derived proximal 15q duplication. *Am. J. Hum. Genet.* **60,** 928–934 (1997).
19. Sebat, J. *et al.* Strong association of *de novo* copy number mutations with autism. *Science* **316,** 445–449 (2007).
This important study found a higher proportion of *de novo* CNVs in simplex autism families than in multiplex autism families, suggesting that the underlying genetic mechanisms of disease may be distinct in these cases.
20. Szatmari, P. *et al.* Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nature Genet.* **39,** 319–328 (2007).
This comprehensive study of linkage and CNVs delineates the importance of genetic and phenotypic heterogeneity in autism, as well as how CNVs can be used to stratify cohorts.
21. Marshall, C. R. *et al.* Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.* **82,** 477–488 (2008).
22. Weiss, L. A. *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* **358,** 667–675 (2008).
23. Kumar, R. A. *et al.* Recurrent 16p11.2 microdeletions in autism. *Hum. Mol. Genet.* **17,** 628–638 (2008).
24. Christian, S. L. *et al.* Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder. *Biol. Psychiatry* **63,** 1111–1117 (2008).
25. Jacquemont, M. L. *et al.* Array-based comparative genomic hybridisation identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders. *J. Med. Genet.* **43,** 843–849 (2006).
26. Morrow, E. M. *et al.* Identifying autism loci and genes by tracing recent shared ancestry. *Science* **321,** 218–223 (2008).
27. Walsh, T. *et al.* Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320,** 539–543 (2008).
28. Xu, B. *et al.* Strong association of *de novo* copy number mutations with sporadic schizophrenia. *Nature Genet.* **40,** 880–885 (2008).
29. The International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455,** 237–241 (2008).
30. Stefansson, H. *et al.* Large recurrent microdeletions associated with schizophrenia. *Nature* **455,** 232–236 (2008).
This novel study exploited rare recurrent CNVs discovered to be under negative selection through population genetic studies and applied this information to identify schizophrenia-susceptibility loci.
31. de Vries, B. B. *et al.* Diagnostic genome profiling in mental retardation. *Am. J. Hum. Genet.* **77,** 606–616 (2005).
32. Sharp, A. J. *et al.* Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nature Genet.* **38,** 1038–1042 (2006).
33. Mefford, H. C. *et al.* Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N. Engl. J. Med.* doi:10.1056/NEJMoa0805384; published online 10 September 2008.
34. Sharp, A. J. *et al.* A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nature Genet.* **40,** 322–328 (2008).
35. Miller, D. T. *et al.* Microdeletion/duplication at 15q13.2q13.3 among individuals with features of autism and other neuropsychiatric disorders. *J. Med. Genet.* doi:10.1136/jmg.2008.059907; published online 19 September 2008.
36. Tabor, H. K. & Cho, M. K. Ethical implications of array comparative genomic hybridization in complex phenotypes: points to consider in research. *Genet. Med.* **9,** 626–631 (2007).
This paper provides a thorough discussion of the relevance of CNV data in complex disease studies, including ethical considerations arising in this new field.
37. Stranger, B. E. *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315,** 848–853 (2007).
38. Durand, C. M. *et al.* Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nature Genet.* **39,** 25–27 (2007).
39. Bruder, C. E. *et al.* Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *Am. J. Hum. Genet.* **82,** 763–771 (2008).
40. Zhao, X. *et al.* A unified genetic theory for sporadic and inherited autism. *Proc. Natl Acad. Sci. USA* **104,** 12831–12836 (2007).
41. Beaudet, A. L. Autism: highly heritable but not inherited. *Nature Med.* **13,** 534–536 (2007).
42. Darilek, S. *et al.* Pre- and postnatal genetic testing by array-comparative genomic hybridization: genetic counseling perspectives. *Genet. Med.* **10,** 13–18 (2008).
43. McCarthy, M. I. *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Rev. Genet.* **9,** 356–369 (2008).
44. Scherer, S. W. *et al.* Challenges and standards in integrating surveys of structural variation. *Nature Genet.* **39,** S7–S15 (2007).
45. De Luca, A. *et al.* Deletions of *NF1* gene and exons detected by multiplex ligation-dependent probe amplification. *J. Med. Genet.* **44,** 800–808 (2007).
46. Grisart, B. *et al.* *NF1* microduplication first clinical report: association with mild mental retardation, early onset of baldness and dental enamel hypoplasia? *Eur. J. Hum. Genet.* **16,** 305–311 (2008).
47. Lee, J. A., Carvalho, C. M. & Lupski, J. R. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* **131,** 1235–1247 (2007).
48. Lachman, H. M. *et al.* Increase in GSK3β gene copy number variation in bipolar disorder. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **144,** 259–265 (2007).
49. Chartier-Harlin, M. C. *et al.* α-Synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet* **364,** 1167–1169 (2004).
50. Ibanez, P. *et al.* Causal relation between α-synuclein gene duplication and familial Parkinson's disease. *Lancet* **364,** 1169–1171 (2004).
51. Shelley, B. P., Robertson, M. M. & Turk, J. An individual with Gilles de la Tourette syndrome and Smith–Magenis microdeletion syndrome: is chromosome 17p11.2 a candidate region for Tourette syndrome putative susceptibility genes? *J. Intellect. Disabil. Res.* **51,** 620–624 (2007).
52. Lawson-Yuen, A., Saldivar, J. S., Sommer, S. & Picker, J. Familial deletion within *NLGN4* associated with autism and Tourette syndrome. *Eur. J. Hum. Genet.* **16,** 614–618 (2008).

Acknowledgements We thank P. Szatmari, R. Weksberg, C. Marshall, L. Feuk and D. Pinto for ideas. Research in our laboratories is supported by The Centre for Applied Genomics, Genome Canada–Ontario Genomics Institute, the Canadian Institutes for Health Research (CIHR), the Canadian Institute for Advanced Research, the McLaughlin Centre for Molecular Medicine, the Canadian Foundation for Innovation, the Ontario Ministry of Research & Innovation, the National Institutes of Health (grant number HD055751, to E.H.C.), Children's Brain Research Foundation, Jean Young and Walden W. Shaw Foundation, Autism Speaks, and the SickKids Foundation. S.W.S. holds the GlaxoSmithKline–CIHR Chair in Genetics and Genomics at the University of Toronto and the Hospital for Sick Children.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence should be addressed to S.W.S. (steve@genet.sickkids.on.ca).

ARTICLES

The cranial endoskeleton of *Tiktaalik roseae*

Jason P. Downs[1], Edward B. Daeschler[1], Farish A. Jenkins Jr[2] & Neil H. Shubin[3,4]

Among the morphological changes that occurred during the 'fish-to-tetrapod' transition was a marked reorganization of the cranial endoskeleton. Details of this transition, including the sequence of character acquisition, have not been evident from the fossil record. Here we describe the braincase, palatoquadrate and branchial skeleton of *Tiktaalik roseae*, the Late Devonian sarcopterygian fish most closely related to tetrapods. Although retaining a primitive configuration in many respects, the cranial endoskeleton of *T. roseae* shares derived features with tetrapods such as a large basal articulation and a flat, horizontally oriented entopterygoid. Other features in *T. roseae*, like the short, straight hyomandibula, show morphology intermediate between the condition observed in more primitive fish and that observed in tetrapods. The combination of characters in *T. roseae* helps to resolve the relative timing of modifications in the cranial endoskeleton. The sequence of modifications suggests changes in head mobility and intracranial kinesis that have ramifications for the origin of vertebrate terrestriality.

The vertebrate transition to land was accompanied by changes in a variety of functional systems including feeding, respiration, support and locomotion. The relative timing of morphological changes associated with these shifts is revealed exclusively through the study of fossil taxa. In recent years, our understanding of vertebrate terrestrialization has been greatly expanded by the description of intermediate forms such as *Panderichthys rhombolepis* from the late Givetian to early Frasnian of Latvia[1,2], *Elpistostege watsoni* from the early Frasnian of Quebec[3,4], and most recently *Tiktaalik roseae* from the early–middle Frasnian of Nunavut[5,6]. These and other taxa have revealed that some of the innovations traditionally associated with terrestriality first appeared in aquatic elpistostegalians. Examples of such features include body-supporting forelimbs and pectoral girdle; a neck supporting a head capable of independent motion; and raised, dorsally positioned eyes on a flattened skull. Recent phylogenetic analyses[5,7,8,9] distribute these features along the tetrapod stem and, as a result, the origin of tetrapods in the Late Devonian (375–363 million years (Myr) ago) is now recognized as a stepwise process of character acquisition, rather than an abrupt shift. Despite this progress, a morphological gap has persisted between the cranial endoskeleton of tetrapodomorph fish (finned tetrapodomorphs) and that of tetrapods (limbed tetrapodomorphs). As a consequence of this apparent morphological disparity, numerous synapomorphies have been used to distinguish the tetrapod cranial endoskeleton. Examples include reorganization of the basicranium via the reorientation of the vestibular fontanelle and loss of the lateral commissure, basicranial fenestra and intracranial joint; reduction and transverse reorientation of the hyomandibula; and changes in the shape and articulations of the palate. Owing to the lack of data from intermediate taxa, the prevailing perception has been that the transition of the cranial endoskeleton to the tetrapod condition was a rapid event[10,11].

T. roseae, with a phylogenetic position between *P. rhombolepis* and *Acanthostega gunnari*[5,8,9], provides an opportunity to fill this gap in our understanding. Multiple specimens of *T. roseae* were prepared to reveal the cranial endoskeleton in exceptional three-dimensional detail. These new data offer insight into the evolution of the cranial endoskeleton during the vertebrate transition into shallow water and amphibious lifestyles.

Description

The braincase of *T. roseae*, like that of other tetrapodomorph fish, is comprised of ethmosphenoidal and otic-occipital components with unfused but complex contact surfaces between them (Figs 1–3; see Supplementary Information for additional photographs). The ethmosphenoid of *T. roseae* possesses ossified nasal capsules and a parasphenoid that extends rostrally to a position between the vomers. The parasphenoid is perforated by a buccohypophyseal foramen (Fig. 3). A large, laterally projecting basipterygoid process carries, at its distal end, a dorsorostrally facing, concave surface for articulation with the palatoquadrate (Figs 2 and 3). In addition, the ethmosphenoid, like that of other finned tetrapodomorphs, has two other articulations with the palate: one is on the postnasal wall and the other is caudodorsal to the basipterygoid process (Fig. 2). *A. gunnari* has only an enlarged basipterygoid process as a point of articulation[12].

The otic-occipital region of the *T. roseae* braincase shares with other tetrapodomorph fish a basicranial fenestra, a lateral commissure, a groove for the jugular vein that leads to a canal through the lateral commissure, and a ventrally directed vestibular fontanelle (Figs 1–3). Several features of the otic-occipital are shared exclusively with that of *P. rhombolepis*, particularly a caudolaterally projecting lateral commissure and a broad basioccipital that gives this part of the braincase a bulbous rather than rectangular appearance in ventral view[10]. The lobate prootic process of *T. roseae* extends rostrally to insert into a recess on the ventral surface of the postorbital (Figs 2 and 3). The prootic process of *A. gunnari* similarly articulates with a recessed facet of the tabular[13]. Although *T. roseae* is the only finned tetrapodomorph known to possess this process, its occurrence in certain taxa, including *P. rhombolepis*, is presently impossible to determine because the palate obscures this region in the available material. *Eusthenopteron foordi* does exhibit 'anterolateral bars' at

[1]Academy of Natural Sciences of Philadelphia, 1900 Benjamin Franklin Parkway, Philadelphia, Pennsylvania 19103, USA. [2]Harvard University, Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, 26 Oxford Street, Cambridge, Massachusetts 02138, USA. [3]University of Chicago, Department of Organismal Biology and Anatomy, 1027 E. 57th Street, Chicago, Illinois 60637, USA. [4]Field Museum of Natural History, Chicago, Illinois 60605, USA.

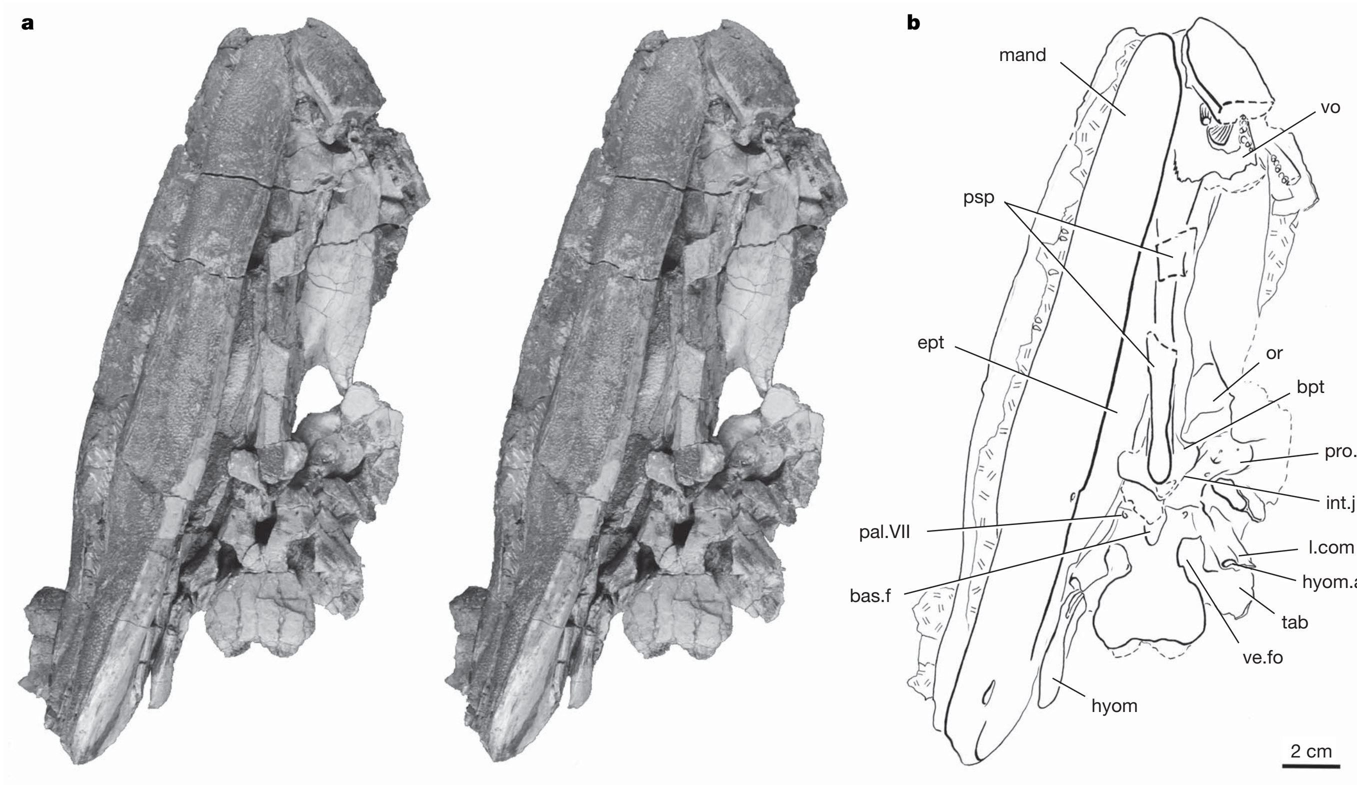


Figure 1 | Nunavut Fossil Vertebrate Collection (NUFV) 110, skull of *Tiktaalik roseae* in ventral view. a, Stereo pair with the left mandible and palatoquadrate removed to expose the braincase. **b**, Interpretive drawing of NUFV 110. Note the right hyomandibula in articulation with the lateral commissure of the braincase. bas.f, basicranial fenestra; bpt, basipterygoid process; ept, entopterygoid; hyom, hyomandibula; hyom.art, articulation with hyomandibula; int.j, intracranial joint; l.com, lateral commissure; mand, mandible; or, orbit; pal.VII, foramen for palatine branch of cranial nerve VII; pro.p, prootic process; psp, parasphenoid; tab, tabular; ve.fo, vestibular fontanelle; vo, vomer.

the rostrolateral corners of the otic-occipital that may represent poorly ossified prootic processes[13].

Comparable to the conditions reported in *P. rhombolepis*[14] and *A. gunnari*[12], the entopterygoid of *T. roseae* is horizontal and shallow (Fig. 2). This stands in stark contrast to the sinuous, dorsally reaching element of basal tetrapodomorphs like *E. foordi*. Accordingly, like that of *A. gunnari*[12], the basal articulation of *T. roseae* lies in the same plane as the rest of the entopterygoid. As in other tetrapodomorph fish, the entopterygoids are separated by the braincase along their entire length (Fig. 1) and a denticulated field covers much of the

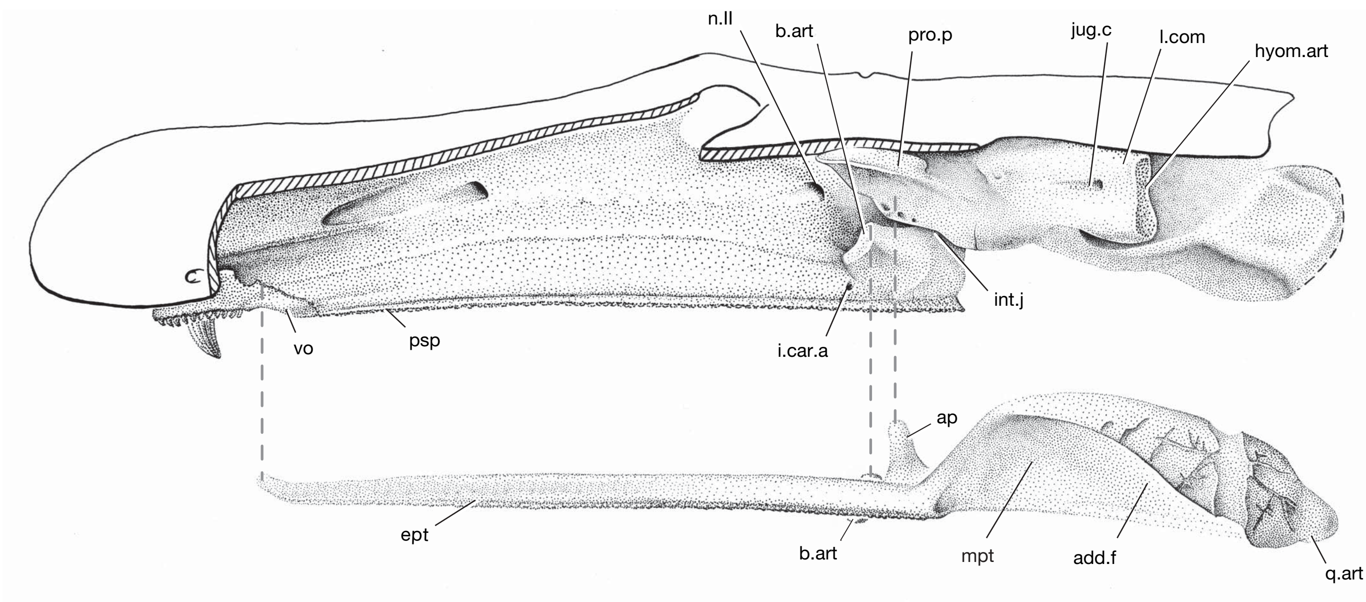


Figure 2 | Reconstruction of the braincase and palatoquadrate of *Tiktaalik roseae* in lateral view. Hashed lines indicate the three articulation sites between braincase and palatoquadrate. add.f, adductor fossa; ap, ascending process; b.art, basal articulation; ept, entopterygoid; hyom.art, articulation with hyomandibula; i.car.a, internal carotid artery; int.j, intracranial joint; jug.c, jugular canal; l.com, lateral commissure; mpt, metapterygoid; n.II, foramen for cranial nerve II; pro.p, prootic process; psp, parasphenoid; q.art, articulation surface of quadrate; vo, vomer.

elements' buccal surfaces. Caudal to the denticulated field, the entopterygoid increases in depth and narrows to a vertically oriented bar at its caudal juncture with the metapterygoid. The metapterygoid is flat, vertical and tall and therefore compares more closely with that of other elpistostegalians[12,14] than with the shallow, wide metapterygoid of basal tetrapodomorphs[15]. Caudal to the basal articulation, a dorsomedially projecting, cylindrical ascending process (columella cranii of ref. 12) terminates against the ethmosphenoid in a position medial to the prootic process (Fig. 2; see Supplementary Information for photographs). The same process in *A. gunnari* is interpreted to meet neither the skull roof nor the braincase[12].

The ossified portion of the hyomandibula of *T. roseae* is short and straight, unlike the boomerang-shaped element of basal finned tetrapodomorphs[15–21,23] (Figs 4 and 5; see Supplementary Information for photographs). The extent to which the hyomandibula continued distally in cartilage cannot be determined. Rather than being steeply inclined like that of *E. foordi*, it is oriented horizontally like that of *P. rhombolepis*[14]. A longitudinal sulcus on the medial surface adjacent to the ventral margin (Fig. 4b) represents the passage for the hyomandibular nerve as in other tetrapodomorph fish[15–23]. Unlike the condition in these taxa, the groove does not lead to a bony fenestra. The hyomandibula of *T. roseae* articulates with the braincase along the distal edge of the lateral commissure. In contrast to the condition in other finned tetrapodomorphs[15–21], however, the hyomandibula articulates with the lateral commissure along a single, strap-shaped surface rather than at separate dorsal and ventral facets (Fig. 4c). As has been proposed for *P. rhombolepis*[14], the hyomandibula of *T. roseae* does not articulate with the palatoquadrate, and cartilaginous elements may have connected the hyomandibula to the ventral hyoid elements. *T. roseae* lacks an ossified operculum and, accordingly, the hyomandibula exhibits no opercular articulation.

Ventral elements of the branchial skeleton are known for *T. roseae*, including two basibranchials; the urohyal; right and left hypohyals and ceratohyals; hypobranchials for the first three arches; and ceratobranchials for four arches (Fig. 6; see Supplementary Information for photograph). As in *Gogonasus andrewsae*[20], *Medoevia lata*[23], *Mandageria fairfaxi*[24] and *E. foordi*[15], the larger, more rostral of the two basibranchials bears articulations for the hypohyals, the hypobranchials of the first two arches, and the second basibranchial. The octagonal shape (in dorsal and ventral views) of basibranchial 1 in *T. roseae* is similar to the octagonal shape of the first basibranchial in *G. andrewsae*[20] and *M. lata*[23], and is quite unlike the longer, more rectangular element of the tristichopterids *M. fairfaxi*[24] and *E. foordi*[15]. Two of three *T. roseae* specimens exhibit dental plate fragments attached to the dorsal surface of this element. Whether basibranchial 1 of *T. roseae* carries one, two, or more dental plate(s) is impossible to determine from the preserved fragments. Basibranchial 2 is an arched (concave ventral) element that appears to carry the articulations for the third arch hypobranchials on its caudolateral margins. The ventral concavity of the second basibranchial and a depression on the ventral face of the first (the ventral basibranchial fossa) accommodate the urohyal, which bears a deep midsagittal crest on its ventral aspect. A site lacking perichondral bone on the medial margin of hypobranchial 3 probably represents the articulation with the fourth branchial arch, a condition observed in dipnomorphs and other finned tetrapodomorphs[24]. *T. roseae* has deeply grooved ceratobranchials like those of *G. andrewsae*[20], *E. foordi*[15] and *A. gunnari*[25]. Typical of gilled osteichthyans, the groove of the ceratobranchial houses the afferent branchial artery[25].

Discussion

The transition from the cranial endoskeleton of finned tetrapodomorphs, such as *E. foordi*, to that of limbed forms, such as *A. gunnari*, entailed a number of morphological changes including a reduction and reorientation of the hyomandibula, reorganization of the basicranium and the articulations between palate and braincase, and alteration of skull architecture including flattening of the head, loss of the intracranial joint, and lengthening of the ethmosphenoid relative to the otic-occipital. *T. roseae*, the sister group of tetrapods, provides new context in which to interpret these alterations as a

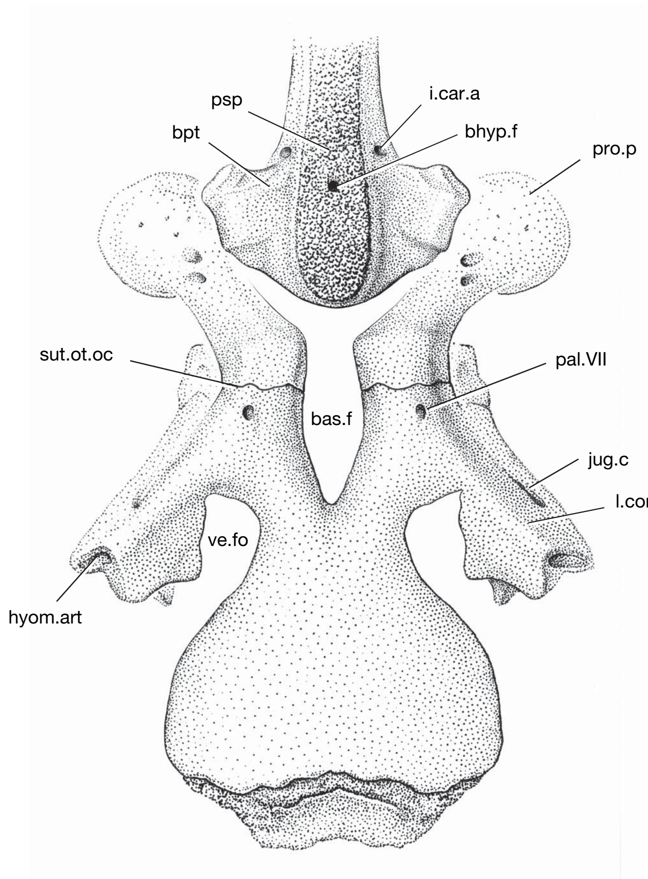


Figure 3 | Reconstruction of the otic-occipital and caudal end of the ethmosphenoid of *Tiktaalik roseae* in ventral view based primarily on NUFV 110. Note that the location and size of the buccohypophyseal foramen and jugular canal are based on NUFV 109 because these structures are not visible in NUFV 110. bas.f, basicranial fenestra; bhyp.f, buccohypophyseal foramen; bpt, basipterygoid process; hyom.art, articulation with hyomandibula; i.car.a, foramen for internal carotid artery; jug.c, jugular canal; l.com, lateral commissure; pal.VII, foramen for palatine branch of cranial nerve VII; pro.p, prootic process; psp, parasphenoid; sut.ot.oc, otic-occipital suture; ve.fo, vestibular fontanelle.

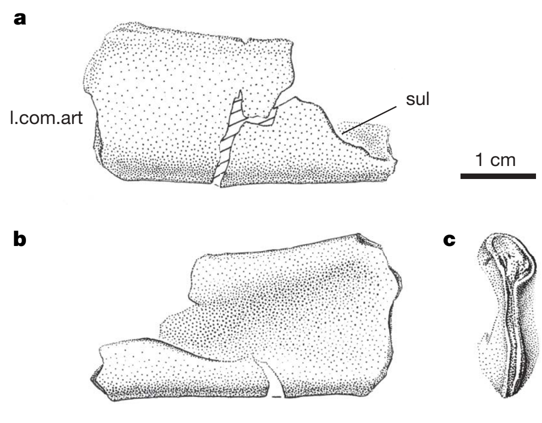


Figure 4 | NUFV 110, left hyomandibula of *Tiktaalik roseae*. a, Lateral view. **b**, Medial view. **c**, Proximal view. Note that the unfinished distal margin is not the result of breakage but rather represents the limit of ossification. l.com.art, articular surface with lateral commissure; sul, medial sulcus.

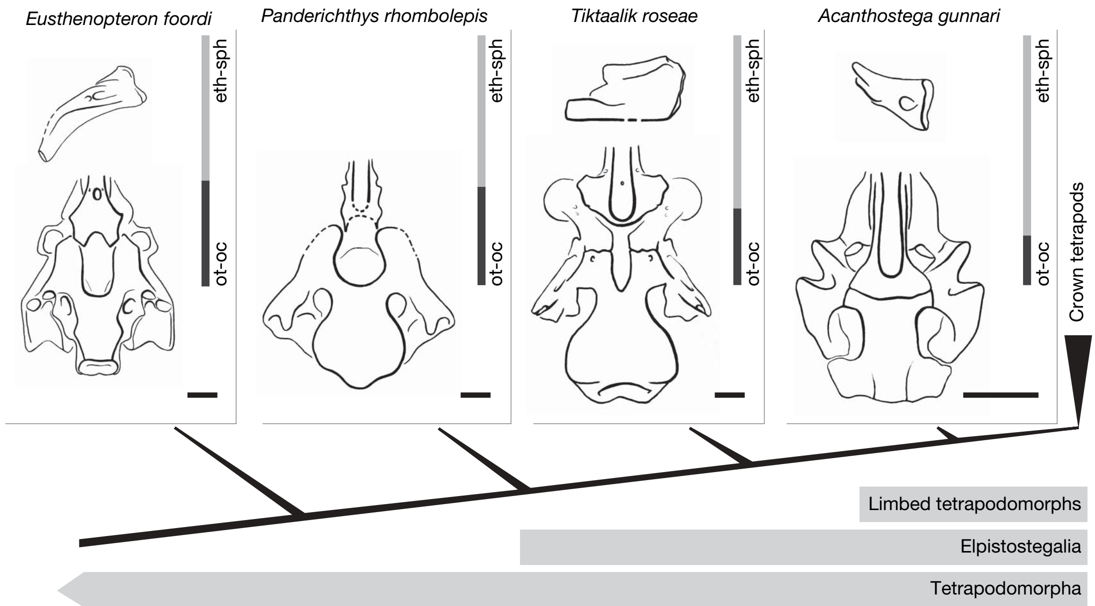


Figure 5 | Cladogram of select tetrapodomorphs. Topology according to ref. 5. Reconstructions of the caudal end of the braincase in ventral view for each of four tetrapodomorph taxa (*Eusthenopteron foordi*, *Panderichthys rhombolepis*, *Tiktaalik roseae* and *Acanthostega gunnari*); reconstructions of the left hyomandibula in medial view (left stapes in caudal view for *A. gunnari*) for taxa in which the element is well known (*E. foordi*, *T. roseae* and *A. gunnari*). The bicoloured bars represent relative length proportions of ethmosphenoid (eth-sph) and otic-occipital (ot-oc) for each of the four represented taxa (*E. foordi*, SMNH (Swedish Museum of Natural History) P.222; *P. rhombolepis*, LDM (Latvian Museum of Natural History) 60/123; *T. roseae*, NUFV 110; *A. gunnari*, based on the reconstruction of ref. 13). *T. roseae* shares with more basal tetrapodomorphs the presence of a buccohypophyseal foramen, ossified nasal capsules, three sites of articulation between the braincase and palate, a basicranial fenestra, a lateral commissure, a ventrally directed vestibular fontanelle, and an unfused endoskeletal intracranial joint. *T. roseae* shares with tetrapods an enlarged basal articulation and a flat, horizontally oriented palatoquadrate. The reduced hyomandibula of *T. roseae* appears intermediate between the condition observed in other finned tetrapodomorphs and that observed in tetrapods. Reconstructions are modified from refs 10, 13, 15 and 31. Scale bars equal 1 cm.

sequence of changes (Fig. 5). Changes in cranial dimensions and reduction of the hyomandibula began before the enlargement of the basipterygoid process[5,7,8]. These changes preceded major alterations in braincase morphology including loss of the lateral commissure and basicranial fenestra, reorientation of the vestibular fontanelle, and loss of the intracranial joint (Fig. 5). The plesiomorphies retained by *T. roseae* reinforce the suggestion that significant changes to the braincase occurred relatively late in the transition to the tetrapod condition[10,11]. The overall sequence of modifications to the cranial endoskeleton suggests changes in head mobility and cranial kinesis that carry important implications for feeding and respiration in Late Devonian marginal aquatic environments.

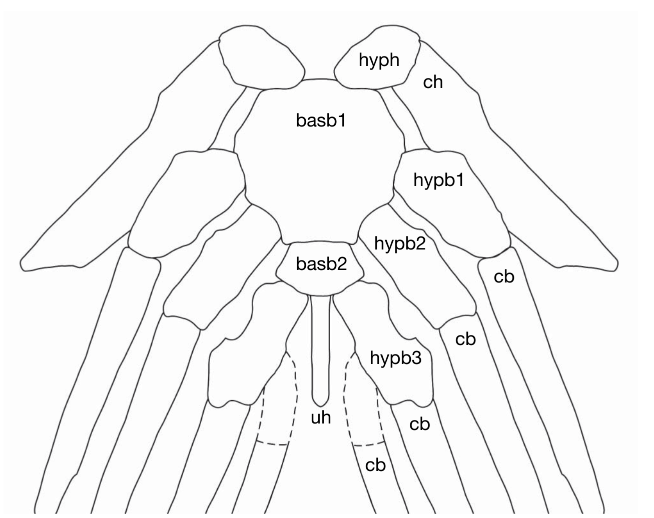


Figure 6 | Schematic diagram of ventral branchial elements of *Tiktaalik roseae* in dorsal view. basb1, basibranchial 1; basb2, basibranchial 2; cb, ceratobranchial; ch, ceratohyal; hyph, hypohyal; hypb1, hypobranchial 1; hypb2, hypobranchial 2; hypb3, hypobranchial 3; uh, urohyal.

A flat and broad skull, as seen in *P. rhombolepis* and more crownward tetrapodomorphs, had structural ramifications for virtually every component of the cranial endoskeleton. In *T. roseae*, as in tetrapods, a diminution of cranial depth is associated with the flattening of the palatoquadrate and the ventral displacement and increase in the size of the basal articulation. These alterations are coincident with the trend towards lengthening the ethmosphenoid relative to the otic-occipital[10,11]. With cranial proportions approaching those of *A. gunnari*, *T. roseae* supports the hypothesis that the general cranial dimensions of tetrapods were attained in tetrapodomorph fish (Fig. 5).

These changes in the shape of the head seem to be related to the reduction and reorientation of the hyomandibula. In the primitive condition, as in *E. foordi*, the bone is boomerang-shaped and ventrally inclined, whereas in elpistostegalians, the hyomandibula is a short, straight, horizontal element (Fig. 5). In *E. foordi*, the hyomandibula articulates with the braincase, palatoquadrate, operculum and branchial skeleton. In *P. rhombolepis*, the ossified portion of the hyomandibula does not articulate with the palatoquadrate and extends from the lateral commissure to a facet on the opercular bone[14]. The stapes (the homologue of the hyomandibula) of *A. gunnari* is oriented transversely, extending laterally from the fenestra vestibuli of the braincase (in part, the homologue of the vestibular fontanelle)[13,26,27]. The hyomandibula of *T. roseae* appears intermediate in size and orientation between those of *P. rhombolepis* and *A. gunnari*. Projecting caudolaterally from its articulation with the lateral commissure, it does not articulate with the palatoquadrate and, concomitant with the lack of an opercular bone, it has no opercular articulation.

With the loss of opercular and extrascapular elements, head mobility is enhanced in *T. roseae* relative to other tetrapodomorph fish. The establishment of a neck has clear ramifications for feeding and

locomotion in shallow water and marginal aquatic settings. Fish move and feed in three-dimensional space and are readily able to orient the body in order to position the mouth towards prey. A neck is advantageous in settings where the body is relatively fixed, as is the case in shallow pools of water or in marginal settings where the body is supported by appendages planted against a substrate.

In the transition to land, physical parameters were altered from a relatively viscous, neutrally buoyant environment to one in which oxygen availability and the effects of gravity are radically different from water. Intracranial kinesis in fish is an adaptation for feeding and respiring in the viscous, fluid medium of water. In tetrapods, intracranial kinesis is lost. Although the head has become more mobile in *T. roseae*, intracranial kinesis is restricted relative to more primitive tetrapodomorphs. Although *T. roseae* retains the bipartite components of the braincase, loss of the dermal intracranial joint, prootic processes that articulate with the postorbitals, and the complexity of the apposing ethmosphenoidal and otic-occipital surfaces obviate the possibility of motion at their contact[28]. In addition, the enlarged size of the basal articulation points to a palate less capable of motion relative to the braincase. Likewise, the loss of the primitive bicondylar joint between hyomandibula and lateral commissure is suggestive of a restricted range of motion. *T. roseae*, then, reveals an intermediate stage in the origin of the consolidated skull of tetrapods.

As the skull becomes a more solid structure, the role of the hyomandibula in respiration is diminished. Primitively, by forming a structural link between the branchial skeleton, operculum, palate and lower jaw, the hyomandibula helps to coordinate their motions[28,29]. The hyomandibula in obligate gill breathers is part of an osseous system that accommodates the stresses associated with branchial and opercular pumping of a relatively viscous medium (water) across the gills. In tetrapods, this function of the hyomandibula is entirely lost, making the element available for an eventual role in hearing[30]. *T. roseae* reveals an intermediate stage in this transition. In *T. roseae*, the reduction and reorientation of the hyomandibula and the consequent loss of its articulations with other components of the cranial endoskeleton suggest a decreasing reliance on water pumping as a mode of respiration.

With appendages able to support the weight of the body, a more consolidated skull, ribs capable of sustaining the trunk under a gravitational load, a respiratory system less reliant on water breathing, a head capable of independent motion, and cranial proportions approaching those of tetrapods, *Tiktaalik* reveals an important stage in the origin of terrestrial vertebrates[5,6]. *T. roseae* probably inhabited the benthos, shallows, and mudflats of freshwater floodplains while supporting itself on a solid substrate. In the Late Devonian, marginal freshwater environments were an important locus of the evolutionary changes that led to the origin of amphibious lifestyles among tetrapodomorphs.

Note added in proof: Reference 9, published after acceptance of this Article, provides relevant details regarding tetrapodomorph phylogeny across the fin–limb transition.

Received 21 November 2007; accepted 23 June 2008.

1. Vorobyeva, E. I. Observations on two rhipidistian fishes from the Upper Devonian of Lode, Latvia. *Zool. J. Linn. Soc.* **70**, 191–201 (1980).
2. Vorobyeva, E. I. & Schultze, H.-P. in *Origins of the Higher Groups of Tetrapods* (eds Schultze, H.-P. & Trueb, L.) 68–109 (Cornell Univ. Press, 1991).
3. Schultze, H.-P. & Arsenault, M. The panderichthyid fish Elpistostege: A close relative of tetrapods? *Palaeontology* **28**, 292–309 (1985).
4. Schultze, H.-P. in *Devonian Fishes and Plants of Miguasha, Quebec, Canada* (eds Schultze, H.-P. & Cloutier, R.) 316–327 (Verlag Dr. Friedrich Pfeil, 1996).
5. Daeschler, E. B., Shubin, N. H. & Jenkins, F. A. Jr. A Devonian tetrapod-like fish and the evolution of the tetrapod body plan. *Nature* **440**, 757–763 (2006).
6. Shubin, N. H., Daeschler, E. B. & Jenkins, F. A. Jr. The pectoral fin of *Tiktaalik roseae* and the origin of the tetrapod limb. *Nature* **440**, 764–771 (2006).
7. Ahlberg, P. E. & Johanson, Z. Osteolepiforms and the ancestry of tetrapods. *Nature* **395**, 792–794 (1998).
8. Long, J. A., Young, G. C., Holland, T., Senden, T. J. & Fitzgerald, E. M. G. An exceptional Devonian fish from Australia sheds light on tetrapod origins. *Nature* **444**, 199–202 (2006).
9. Ahlberg, P. E., Clack, J. A., Lukševičs, E., Blom, H. & Zupinš, I. *Ventastega curonica* and the origin of tetrapod morphology. *Nature* **453**, 1199–1204 (2008).
10. Ahlberg, P. E., Clack, J. A. & Lukševičs, E. Rapid braincase evolution between *Panderichthys* and the earliest tetrapods. *Nature* **381**, 61–64 (1996).
11. Clack, J. A. The Origin of Tetrapods. in *Amphibian Biology* Vol. 4 (eds Heatwole, H. & Carroll, R.) 979–1029 (Surrey Beatty and Sons, 2000).
12. Clack, J. A. *Acanthostega gunnari*, a Devonian tetrapod from Greenland; the snout, palate, and ventral parts of the braincase, with a discussion of their significance. *Meddelelser Grønland Geosci.* **31**, 1–24 (1994).
13. Clack, J. A. The neurocranium of *Acanthostega gunnari* Jarvik and the evolution of the otic region in tetrapods. *Zool. J. Linn. Soc.* **122**, 61–97 (1998).
14. Brazeau, M. D. & Ahlberg, P. E. Tetrapod-like middle ear architecture in a Devonian fish. *Nature* **439**, 318–321 (2006).
15. Jarvik, E. *Basic Structure and Evolution of Vertebrates* Vol. 1 (Academic, 1980).
16. Smithson, T. R. & Thomson, K. S. The hyomandibular of *Eusthenopteron foordi* Whiteaves (Pisces: Crossopterygii) and the early evolution of the tetrapod stapes. *Zool. J. Linn. Soc.* **74**, 93–103 (1982).
17. Romer, A. S. The braincase of the Carboniferous crossopterygian *Megalichthys nitidus*. *Bull. Mus. Comp. Zool.* **82**, 1–73 (1937).
18. Romer, A. S. Notes on the crossopterygian hyomandibular and braincase. *J. Morphol.* **69**, 141–180 (1941).
19. Fox, R. C., Campbell, K. S. W., Barwick, R. E. & Long, J. A. A new osteolepiform from the Lower Carboniferous Raymond Formation, Drummond Basin, Queensland. *Mem. Queensl. Mus.* **38**, 97–221 (1995).
20. Long, J. A., Barwick, R. E. & Campbell, K. S. W. Osteology and functional morphology of the osteolepiform fish *Gogonasus andrewsae* Long, 1985, from the Upper Devonian Gogo Formation, Western Australia. *Rec. West. Aus. Mus.* **53** (suppl.), 1–89 (1997).
21. Johanson, Z., Ahlberg, P. & Ritchie, A. The braincase and palate of the tetrapodomorph sarcopterygian *Mandageria fairfaxi*: morphological variability near the fish-tetrapod transition. *Palaeontology* **46**, 271–293 (2003).
22. Thomson, K. S. The endocranium and associated structures in the Middle Devonian rhipidistian fish *Osteolepis*. *Proc. Linn. Soc. Lond. Zool.* **176**, 181–195 (1965).
23. Lebedev, O. A. Morphology of a new osteolepidid fish from Russia. *Bull. Mus. Natl. Hist. Nat. Paris* **17**, 287–341 (1995).
24. Johanson, Z. & Ahlberg, P. E. A new tristichopterid (Osteolepiformes: Sarcopterygii) from the Mandagery Sandstone (Late Devonian, Famennian) near Canowindra, NSW, Australia. *Trans. R. Soc. Edinb.* **88**, 39–68 (1997).
25. Coates, M. I. & Clack, J. Fish-like gills and breathing in the earliest known tetrapod. *Nature* **352**, 234–236 (1991).
26. Clack, J. A. Earliest known tetrapod braincase and the evolution of the stapes and fenestra ovalis. *Nature* **369**, 392–394 (1994).
27. Clack, J. A. Discovery of the earliest-known tetrapod stapes. *Nature* **342**, 425–427 (1989).
28. Thomson, K. S. T. Mechanisms of intracranial kinetics in fossil rhipidistian fishes (Crosspoterygii) and their relatives. *Zool. J. Linn. Soc.* **46**, 223–253 (1967).
29. Carroll, R. L. in *The Terrestrial Environment and the Origin of Land Vertebrates* (ed. Panchen, A. L.) 293–317 (Academic, 1980).
30. Clack, J. A. Pattern and processes in the early evolution of the tetrapod ear. *J. Neurobiol.* **53**, 251–264 (2002).
31. Clack, J. A. in *The Evolutionary Biology of Hearing* (eds Webster, D. B., Ray, R. R. & Popper, A. N.) 405–420 (Springer, 1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The illustrations are the work of K. Monoyios. Specimen preparation was performed by C. F. Mullison. This research was made possible by permits from the Nunavut Ministry of Culture, Languages, Elders and Youth, the Grise Fiord Hamlet and HTA, with logistical support from the Polar Continental Shelf Project and collections support from the Canadian Museum of Nature. M. Brazeau improved interpretation of *P. rhombolepis* and offered input. Assistance in the field (1999–2006) was provided by W. Amaral, B. Atagootak, J. Conrad, M. Davis, S. Gatesy, A. Gillis, B. Kilbourne, S. Madsen, K. Middleton, J. Miller, K. Monoyios, C. Schaff, M. Shapiro, R. Shearman and C. Sullivan. This research was supported by an anonymous donor, the Academy of Natural Sciences of Philadelphia, the Putnam Expeditionary Fund (Harvard University), the University of Chicago, the National Science Foundation, and the National Geographic Society Committee for Research and Exploration.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.P.D. (downs@ansp.org) or N.H.S. (nshubin@uchicago.edu).

 Vol 455|16 October 2008|doi:10.1038/nature07261

ARTICLES

Identification of *ALK* as a major familial neuroblastoma predisposition gene

Yaël P. Mossé[1], Marci Laudenslager[1], Luca Longo[2], Kristina A. Cole[1], Andrew Wood[1], Edward F. Attiyeh[1], Michael J. Laquaglia[1], Rachel Sennett[1], Jill E. Lynch[1], Patrizia Perri[2,3], Geneviève Laureys[4], Frank Speleman[4], Cecilia Kim[5], Cuiping Hou[1,5], Hakon Hakonarson[5,7], Ali Torkamani[6], Nicholas J. Schork[6], Garrett M. Brodeur[1], Gian P. Tonini[2], Eric Rappaport[1], Marcella Devoto[7,8] & John M. Maris[1,9]

Neuroblastoma is a childhood cancer that can be inherited, but the genetic aetiology is largely unknown. Here we show that germline mutations in the anaplastic lymphoma kinase (*ALK*) gene explain most hereditary neuroblastomas, and that activating mutations can also be somatically acquired. We first identified a significant linkage signal at chromosome bands 2p23–24 using a whole-genome scan in neuroblastoma pedigrees. Resequencing of regional candidate genes identified three separate germline missense mutations in the tyrosine kinase domain of *ALK* that segregated with the disease in eight separate families. Resequencing in 194 high-risk neuroblastoma samples showed somatically acquired mutations in the tyrosine kinase domain in 12.4% of samples. Nine of the ten mutations map to critical regions of the kinase domain and were predicted, with high probability, to be oncogenic drivers. Mutations resulted in constitutive phosphorylation, and targeted knockdown of *ALK* messenger RNA resulted in profound inhibition of growth in all cell lines harbouring mutant or amplified *ALK*, as well as in two out of six wild-type cell lines for *ALK*. Our results demonstrate that heritable mutations of *ALK* are the main cause of familial neuroblastoma, and that germline or acquired activation of this cell-surface kinase is a tractable therapeutic target for this lethal paediatric malignancy.

Neuroblastoma is a cancer of early childhood that arises from the developing autonomic nervous system. It is the most common malignancy diagnosed in the first year of life and shows a wide range of clinical phenotypes, with a few patients having tumours that regress spontaneously, whereas most patients have aggressive metastatic disease[1]. These latter neuroblastoma cases have survival probabilities of less then 40% despite intensive chemoradiotherapy, and the disease continues to account for 15% of childhood cancer mortality[1,2]. Tumours from patients with an aggressive phenotype often show amplification of the *MYCN* oncogene[3], and/or deletions of chromosome arms 1p and 11q (ref. 4). However, because *MYCN* is so aberrantly dysregulated, and no putative tumour suppressor gene at 1p and 11q has been shown to harbour inactivating mutations in more than a small percentage of cases, no tractable molecular target approaches at present exist for this disease.

Like most human cancers, a small subset of neuroblastoma cases is inherited in an autosomal dominant manner[5–7]. A family history of the disease is found in about 1–2% of newly diagnosed cases, with a standardized incidence ratio of 9.7 for siblings of index cases[8]. Neuroblastoma pedigrees show notable heterogeneity in the type of tumours that arise, with both benign and malignant forms occurring in the same family[9]. Familial neuroblastoma patients differ from those with sporadic disease in that they are diagnosed at an earlier age and/or with several primary tumours, clinical characteristics that are hallmarks of cancer predisposition syndromes. Because of the lethality of the condition before reproductive age, previous genetic linkage scans have been underpowered and results have been difficult to replicate[10–12]. Notably, neuroblastoma can occur with a spectrum of disorders related to abnormal development of neural-crest-derived tissues, including central congenital hypoventilation syndrome and Hirschsprung's disease. Missense or nonsense mutations in *PHOX2B*, a homeobox gene that is a master regulator of normal autonomic nervous system development, were recently shown to predispose to this rare field defect of the sympathoadrenal lineage tissues[13–15]. However, *PHOX2B* mutations explain only a small subset of hereditary neuroblastoma, are almost exclusive to cases with associated disorders of neural-crest-derived tissues, and are not somatically acquired in tumours[16,17], leaving the genetic aetiology for most familial neuroblastoma cases unknown.

Identification of germline *ALK* mutations

To identify the location of a hereditary neuroblastoma predisposition gene, we performed a genome-wide scan for linkage at ~6,000 single nucleotide polymorphisms (SNPs) in 20 neuroblastoma families. Because of the rarity of the condition, the genome-wide scan included pedigrees with varying degrees of confidence of actual heritability. Eight families had three or more affected individuals of close relation (high confidence), six families consisted of only two individuals of first-degree relation (moderate confidence), and six other families also only had two affected individuals, but of a more distant relationship (low confidence). We discovered a significant linkage signal at chromosome 2p, with a maximum non-parametric

[1]Division of Oncology and Center for Childhood Cancer Research, Children's Hospital of Philadelphia, Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA. [2]Translational Pediatric Oncology, National Institute for Cancer Research and Italian Neuroblastoma Foundation, National Institute for Cancer Research, Genoa 16132, Italy. [3]Advanced Biotechnology Center, Genoa 16132, Italy. [4]Center for Medical Genetics, Ghent University Hospital, Gent B-9000, Belgium. [5]The Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. [6]Scripps Genomic Medicine and the Scripps Research Institute, La Jolla, California 92037, USA. [7]Division of Genetics, The Children's Hospital of Philadelphia, Department of Pediatrics and CCEB, Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA. [8]University of Rome "La Sapienza", Department of Experimental Medicine, Rome 00161, Italy. [9]Abramson Family Cancer Research Institute, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA.

likelihood ratio (lod) score of 4.23 at rs1344063 in 18 of the families (two were excluded owing to insufficient DNA). This refined a region previously reported for one of the pedigrees studied here[10]. By mapping informative recombination events, we defined a predisposition locus at chromosome bands 2p23–p24 delimited by SNPs rs1862110 and rs2008535 with 104 genes including the known neuroblastoma oncogene, *MYCN*[3,18], and the *ALK* oncogene located 13.2 megabases (Mb) centromeric. Despite previous work showing that forced overexpression of *MYCN* to the murine neural crest causes neuroblastoma[18], resequencing of the *MYCN* coding region and 18 kilobases (kb) of surrounding genomic DNA in probands from each linked family showed no disease-causal sequence variations.

We next focused on the *ALK* gene because our group and others had previously identified it as a potential oncogene in neuroblastoma through somatically acquired amplification of the genomic locus[19,20]. In addition, oncogenic fusion proteins leading to constitutive activation of the *ALK* kinase domain occur in many human cancers including anaplastic large cell lymphoma[21], inflammatory myofibroblastic tumours[22], squamous cell carcinomas[23] and non-small-cell lung cancers[24,25]. Resequencing of the 29 *ALK* coding exons identified three separate single base substitutions in the *ALK* tyrosine kinase domain in eight of the probands screened (Fig. 1 and Table 1). These DNA sequence alterations were not present in either single nucleotide polymorphism (dbSNP; http://www.ncbi.nlm.nih.gov/projects/SNP/) or somatic mutation (COSMIC; http://www.sanger.ac.uk/genetics/CGP/cosmic/) databases, and were not detected in direct sequencing of the *ALK* tyrosine kinase domain in 218 normal control alleles. Each substitution was subsequently shown to segregate with the disease in each family (Fig. 1). The sequence variation in family FNB12 (R1275Q) seems to have been acquired *de novo* in the affected father, and non-paternity was excluded by analysis of inheritance of genotypes in this pedigree. There are several asymptomatic obligate carriers identified (FNB2, FNB13, FNB32, FNB52 and FNB56), suggesting that the incomplete penetrance of this disease may be due to lack of the acquisition of a second genetic event such as amplification of the mutant allele, or alternatively, spontaneous regression following malignant transformation in at least a subset of cases. Notable is the very large multiplex family (FNB52), with discordance in twins and several unaffected carriers, which segregates a unique germline mutation (G1128A).

ALK sequence variations occurred only in the families with high or moderate degrees of confidence for harbouring a predisposing allele. Six of the eight families with three or more affected individuals had *ALK* missense alterations. The two families that did not have *ALK* sequence alterations identified were each shown to harbour mutations in the sympathoadrenal lineage-specific *PHOX2B* neurodevelopmental gene[14,16]. Two of the six families consisting of only two affected individuals, but of first-degree relation, had *ALK* sequence variations. Each of these families carried the R1275Q alteration, and in FNB12 we showed that the mutation arose *de novo* in the affected father, whereas in FNB56 the alteration was inherited from an unaffected father (Fig. 1). None of the six families with two distant relations affected with neuroblastoma showed *ALK* alterations, suggesting that the occurrence of a further case of this relatively rare disease in an extended family member was probably a chance occurrence. Because there are several families who share identical mutations, we looked to see if these families shared a common haplotype around the *ALK* gene and showed that the affected individuals with the same mutations did not share haplotypes, arguing against a founder effect.

Because *ALK* functions as an oncogene in other human cancers, we predicted that the sequence variations discovered in the neuroblastoma pedigrees would result in constitutive activation. We therefore used a support vector machine-based statistical classifier to map the putative mutations and determine the probability that they would act as drivers of an oncogenic process[26,27]. Each of the germline alterations occurred at regions of the *ALK* kinase domain that have been shown to be principal targets for cancer driver mutations in other

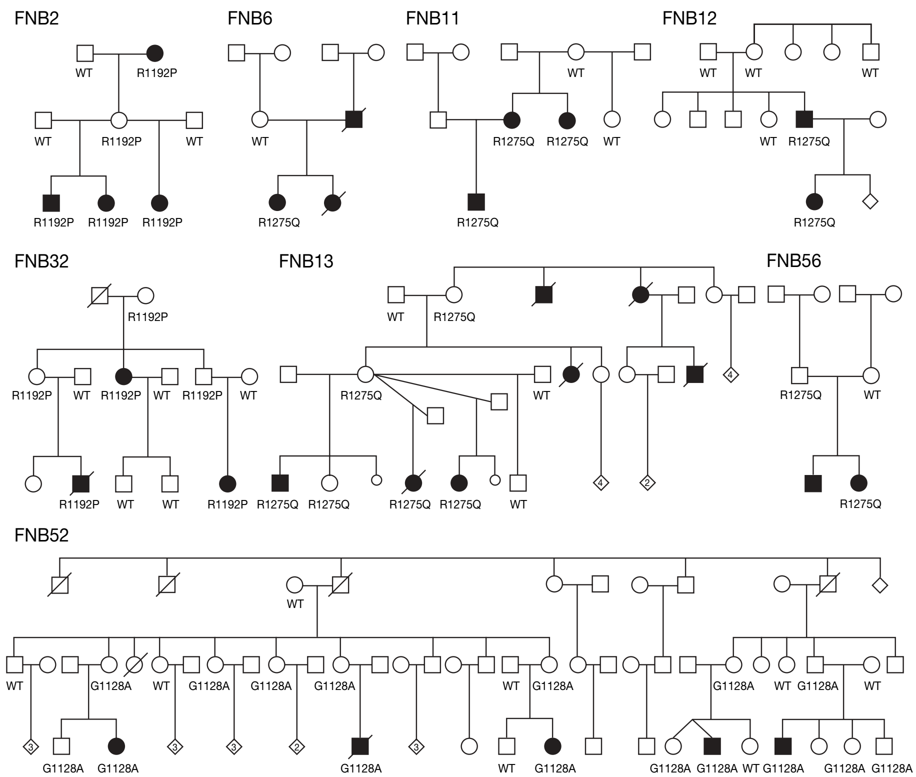


Figure 1 | Eight neuroblastoma pedigrees with *ALK* mutations. All family members with DNA available for genotyping are indicated by either wild-type (WT) for *ALK*, or by mutation in the *ALK* tyrosine kinase domain (R1192P, R1275Q, G1128A). Individuals affected by neuroblastoma are indicated by a filled symbol. The numbers inside the small diamonds indicate the number of other children, the lines through the symbols indicate that the person is deceased, and the smaller circles represent a miscarriage.

Table 1 | *ALK* mutations in neuroblastoma

Mutation	cDNA variation	Type/frequency	Region*	Probability activating mutation†
G1128A	c.3383G > C	Germ line (1/8)	P loop	0.95
R1192P	c.3575G > C	Germ line (2/8)	β4 strand	0.96
R1275Q	c.3824G > A	Germ line (5/8) Somatic (8/24)	Activation loop	0.91
D1091N	c.3271G > A	Somatic (1/24)	N terminus	0.29
M1166R	c.3497T > G	Somatic (1/24)	C helix	0.79
I1171N	c.3512T > A	Somatic (2/24)	C helix	0.85
F1174I	c.3520T > A	Somatic (1/24)	End of C helix	0.92
F1174L	c.3522C > A	Somatic (8/24)	End of C helix	0.96
F1245C	c.3734T > G	Somatic (1/24)	Catalytic loop	0.94
F1245V	c.3733T > G	Somatic (1/24)	Catalytic loop	0.91
I1250T	c.3749T > C	Somatic (1/24)	Catalytic loop	0.87

* The region in which the codon alteration occurs is indicated. Note that the D1091N mutation is immediately adjacent to the tyrosine kinase domain.
† The probability that the amino acid alteration results in oncogenic activation on the basis of previously described methods[26].

oncogenic kinases (Table 1 and Fig. 2). The R1275Q mutation was present in the germline DNA of affected individuals from five pedigrees (Fig. 1), and falls within the kinase activation loop in a region strongly associated with activating mutations in many different protein kinases, such as *BRAF*[28]. This amino acid substitution results in an electropositive residue being replaced by a more electronegative one, possibly mimicking activating phosphorylation events. The R1192P mutation occurred at the beginning of the β4 strand of the kinase domain, and although it is predicted to be a driver mutation with high confidence (Table 1), the mechanism for activation is not yet clear[27]. The G1128A was seen only in the large pedigree with affected individuals in a single generation. The variation falls at the third glycine of the glycine loop, and identical mutations of this glycine residue to alanine in *BRAF* have been shown to increase kinase activity[29].

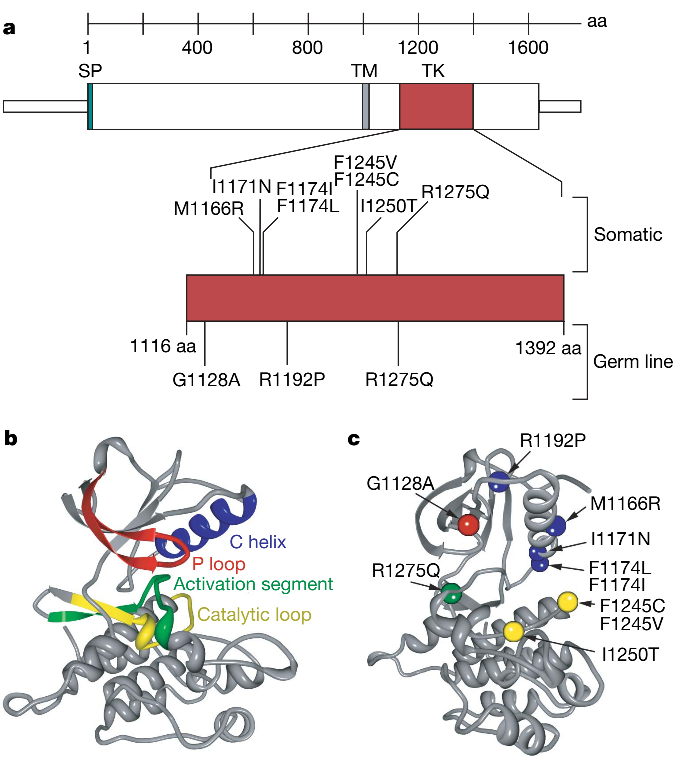


Figure 2 | Germline and somatic *ALK* mutations. **a**, Schematic diagram indicating protein structure of *ALK*, with mutations discovered in constitutional DNAs of familial cases (germ line) and primary tumours from sporadic cases (somatic) indicated. All but one sequence alteration mapped to the tyrosine kinase (TK) domain (D1091N was just amino-terminal and is not indicated here). Of the three germline mutations discovered, only the R1275Q was found in the tumour DNA samples. Conversely, the I1250T mutation discovered in the tumour set was also present in the matched germline DNA of that patient, whereas all of the other mutations studied here were somatically acquired. aa, amino acids; SP, signal peptide; TM, transmembrane. **b**, Homology model of wild-type *ALK* with each main subdomain indicated[26,27]. **c**, *ALK* mutations mapped onto homology model (different orientation from **b** to show all mutations), with colours indicating the subdomain (identified in **b**) in which the mutation resides (for example, the R1275Q mutation falls within the activation segment, indicated in green).

Identification of somatic *ALK* mutations

Having shown that heritable mutations in the *ALK* tyrosine kinase domain are associated with a highly penetrant predisposition to develop neuroblastoma, we next sought to determine whether *ALK* activation might also be somatically acquired. We examined a representative set of 491 sporadically occurring primary neuroblastoma samples acquired from children at the time of diagnosis on a 550K SNP-based microarray to assess for genome-wide copy number alterations. A total of 112 cases (22.8%) showed unbalanced gain of a large genomic region at 2p including the *ALK* locus (partial trisomy), and an additional 16 cases (3.3%) showed high-level focal amplification of *ALK* (Fig. 3). Each of the high-level amplifications co-occurred with *MYCN* amplification and/or other regions at 2p, except one case with an *ALK* amplicon only. The presence of aberrant *ALK* copy number status (gain or amplification) was highly associated with an aggressive clinical phenotype such as metastasis at diagnosis ($P < 0.0001$) and death from disease ($P = 0.0003$).

Because of the association of *ALK* gain and amplification with high-risk disease features, we next examined a subset of 167 tumour samples from high-risk patients, and 27 human neuroblastoma-derived cell lines (all from high-risk patients), for sequence alterations in the *ALK* tyrosine kinase domain. A total of 14 of the 167 tumour (8.4%) and 10 of the 27 cell line (35.7%) samples showed single base substitutions consistent with activating mutations (Fig. 2). Eight separate single base substitutions were identified, with the R1275Q mutation being the only mutation also seen in the germline DNA of the families studied. Again, none of the sequence variations discovered here was in SNP databases or was identified in our resequencing of the *ALK* tyrosine kinase domain in 109 control subjects (218 alleles). Mutations were equally distributed between cases with and without *MYCN* amplification. Only one case had a co-occurrence of an *ALK* mutation (F1174L) and genomic amplification of the *ALK* locus, and in this case the mutated allele was amplified (data not shown). Germline DNA was available for 9 out of 14 patients with *ALK* mutations, and in one of these cases the sequence alteration (I1250T) was also present in the germ line, suggesting a hereditary predisposition that may or may not be *de novo* in this case.

Using the same statistical classifier used for the germline mutations, we showed that all but one of the sequence variations discovered in the tumour tissues were predicted to be activating mutations (Table 1), and the one that showed a low probability (D1091N) was outside the core kinase domain. Most of the somatically acquired mutations fell into either the catalytic loop or the C-helix kinase domains, both frequent sites for oncogenic activating mutations

(Fig. 2). Catalytic loop mutants, especially I1250T, may promote oncogenesis by altering substrate binding or, most probably, by altering packing of the HRD and DFG motifs towards an activated conformation[30]. The mutations observed in the *ALK* C-helix domain occurred at positions within the kinase domain previously found to be mutated in other tumours. I1171N falls at an equivalent weakly oncogenic position in *MET* (M1149T)[31], and the M1166R, F1174I and F1174L mutants fall at equivalent positions mutated in *ERBB2* (D769, V777) and *EGFR* (D761, V769)[32–34].

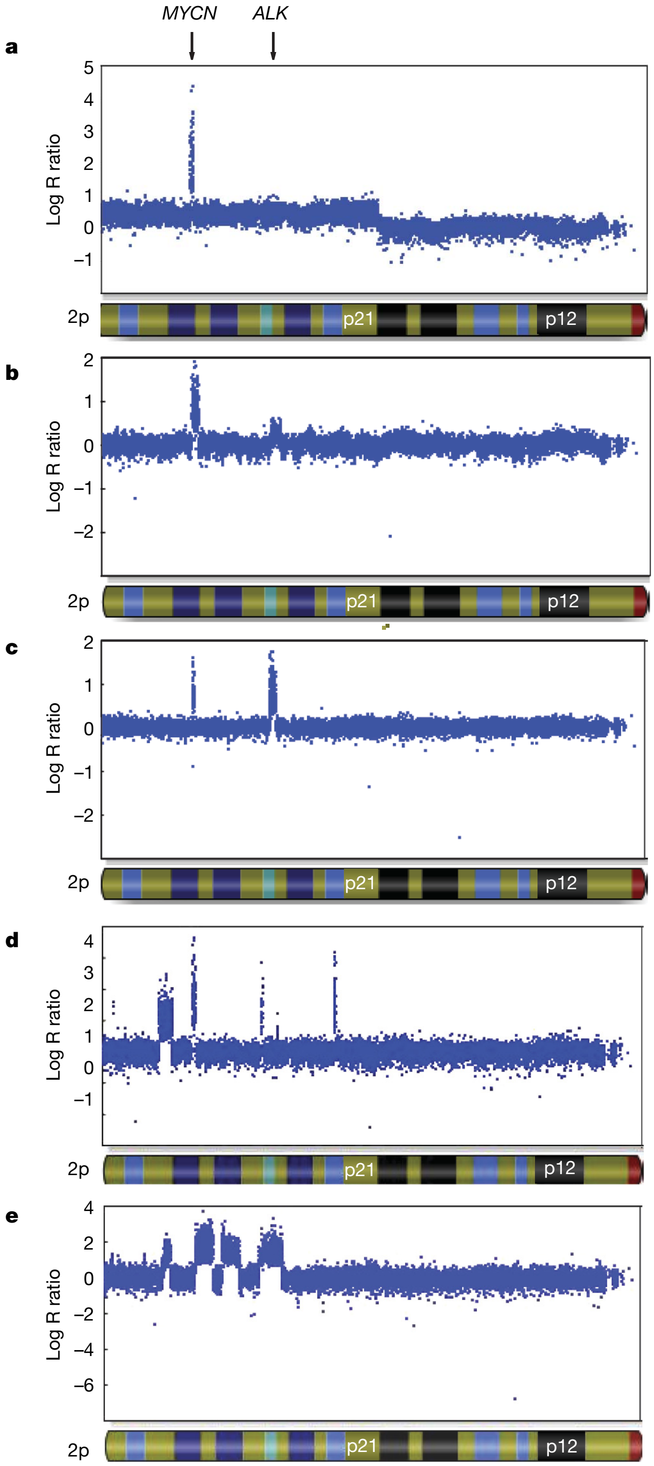


Figure 3 | Representative *ALK* copy number alterations in five neuroblastoma primary tumours. Hybridization intensity reflecting copy number for all SNPs along chromosome 2p in five primary tumours from patients with sporadically occurring disease is shown, and represented on a logarithmic scale. *MYCN* amplification is present in all tumours. **a**, Regional gain (trisomy) of chromosome 2p, including the *ALK* locus. **b**, Focal gain of the *ALK* locus. **c**, Focal amplification of the *ALK* locus. **d**, **e**, Complex rearrangement of the 2p locus, showing various focal amplicons, including *MYCN* and *ALK*.

Functional consequences of *ALK* mutations

We have previously shown that *ALK* is differentially expressed in human primary neuroblastomas with higher expression generally seen in the most aggressive subset of tumours[35]. Using quantitative polymerase chain reaction with reverse transcription (RT–PCR) we confirmed that *ALK* is highly expressed in all but one of the 20 human neuroblastoma-derived cell lines. *ALK* expression was higher in neuroblastoma cells compared to developing fetal brain, and the cell lines harbouring *ALK* mutations ($n = 6$) expressed the mRNA at a significantly higher copy number than the *ALK* wild-type cell lines ($n = 14$, Fig. 4a). Analysis of protein lysates from a panel of neuroblastoma cell lines showed constitutive phosphorylation of the tyrosine residue at codon 1604 in each of the cell lines harbouring mutations, with weak phosphostaining in two wild-type cell lines (Fig. 4b).

To determine whether *ALK* activation by mutation and/or amplification is functionally relevant in models of high-risk neuroblastoma—and thus might offer a tractable therapeutic target—we examined the consequences of disrupting *ALK* signalling by mRNA knockdown. We transiently transfected short interfering RNAs (siRNAs) directed against *ALK* (Dharmacon) into ten neuroblastoma cell lines and screened for inhibition of substrate adherent growth. We demonstrated knockdown of the mRNA and protein in all cell lines studied, but showed differential effects on cellular proliferation (Fig. 5a–l). Each of the cells harbouring *ALK* mutation or amplification showed profound inhibition of proliferation to *ALK* knockdown. In addition, two of the six *ALK* wild-type cell lines showed significant inhibition of growth with *ALK* knockdown and each of these had shown weak evidence for phosphorylation at

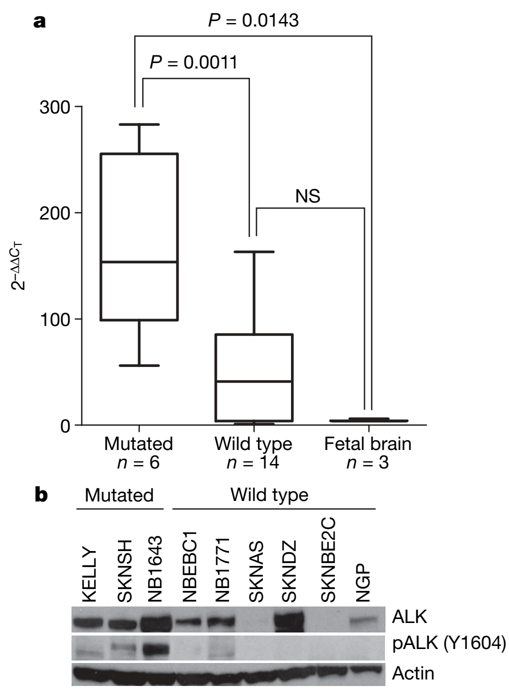


Figure 4 | *ALK* is highly expressed and the kinase is phosphorylated in neuroblastoma cell lines harbouring activating mutations. **a**, Relative *ALK* expression of neuroblastoma cell lines and fetal brain is shown and was determined using the $2^{-\Delta\Delta C_T}$ method[50]. Statistical significance was determined by an unpaired *t*-test. **b**, Immunoblots showing differential ALK expression in neuroblastoma cell lines with phosphorylation of the tyrosine 1604 codon restricted to cell lines with mutations (the wild-type lines NBEBC1 and NB1771 show faint phosphostaining). NS, not significant.

tyrosine 1604 (Fig. 4b), indicating that an alternative mechanism may have resulted in *ALK* kinase activation in these two cell lines.

Discussion

Knudson and Strong predicted that neuroblastoma, like the analogous embryonal cancer retinoblastoma, would follow a two-hit model explaining hereditary and sporadic cases[5]. This model has proved to be correct for most childhood and adult hereditary cancers, and the susceptibility genes are typically tumour suppressors in which the two hits are the sequential inactivation of both alleles. The discovery of heritable mutations in oncogenes as the aetiology of multiple endocrine neoplasia type 2 cancers (*RET*), papillary renal carcinoma (*MET*) and gastrointestinal stromal tumours (*KIT*) challenged this model, but it is now clear that somatically acquired duplication or amplification of the mutant allele provides the second hit[36]. We have shown that heritable mutations in *ALK* are the cause of most hereditary neuroblastoma cases, providing, to our knowledge, the first example of a paediatric cancer arising because of mutations in an oncogene. Together with our recent report that common variations at chromosome band 6p22 predispose to the development of sporadic neuroblastoma[37], the genetic aetiology of this disease is now being defined. The discovery of highly penetrant, heritable *ALK* mutations as the cause of hereditary neuroblastoma is of immediate relevance to patients with a family history; screening with non-invasive techniques such as ultrasonography and measurement of urinary catecholamine metabolites should probably be implemented for unaffected children carrying an *ALK* mutation. In addition, continuing efforts to characterize the full spectrum of mutations in the entire *ALK* coding sequence, as well as determining the frequency of mutations across all neuroblastoma disease subsets, will be required to formulate genetic screening recommendations for newly diagnosed neuroblastoma patients with or without a family history.

ALK is an orphan tyrosine kinase transmembrane receptor with homology to neurotrophin receptors and the *MET* oncogene. Its expression is restricted to the developing nervous system with a postulated role in participating in the regulation of neuronal differentiation[38]. It is now clear that many human cancers activate *ALK* signalling by creating unique oncogenic fusions of *ALK* with a variety of partners through chromosomal translocation events[39]. Previous work had shown that a substantial percentage of human-derived neuroblastoma cell lines express *ALK* transcripts and ALK protein[40], but no definitive role for this oncogene had been proven[19,41–43]. *ALK* was recently identified as a molecular target in neuroblastoma by a screen of human cancer cell lines with pharmacological antagonists of the *ALK* kinase domain[44]. Our current report provides evidence for oncogenic activation of *ALK* by mutation of the kinase domain, and these data provide the genetic basis for the observation of sensitization to *ALK* kinase inhibition. In addition, our discoveries in neuroblastoma may lead to future resequencing efforts in other malignancies, especially those in which oncogenic fusion proteins have recently been discovered. The data presented here clearly establish *ALK* as a critical neuroblastoma oncogene and should increase efforts to identify the ligand for this receptor and to determine whether *ALK*-mediated signalling can be activated by mechanisms other than direct mutation and/or amplification of *ALK* alleles. Furthermore, receptor tyrosine kinases provide tractable targets for pharmacological inhibition, and this work should provide the impetus for developing therapeutic strategies aimed at inhibiting *ALK*-mediated signalling.

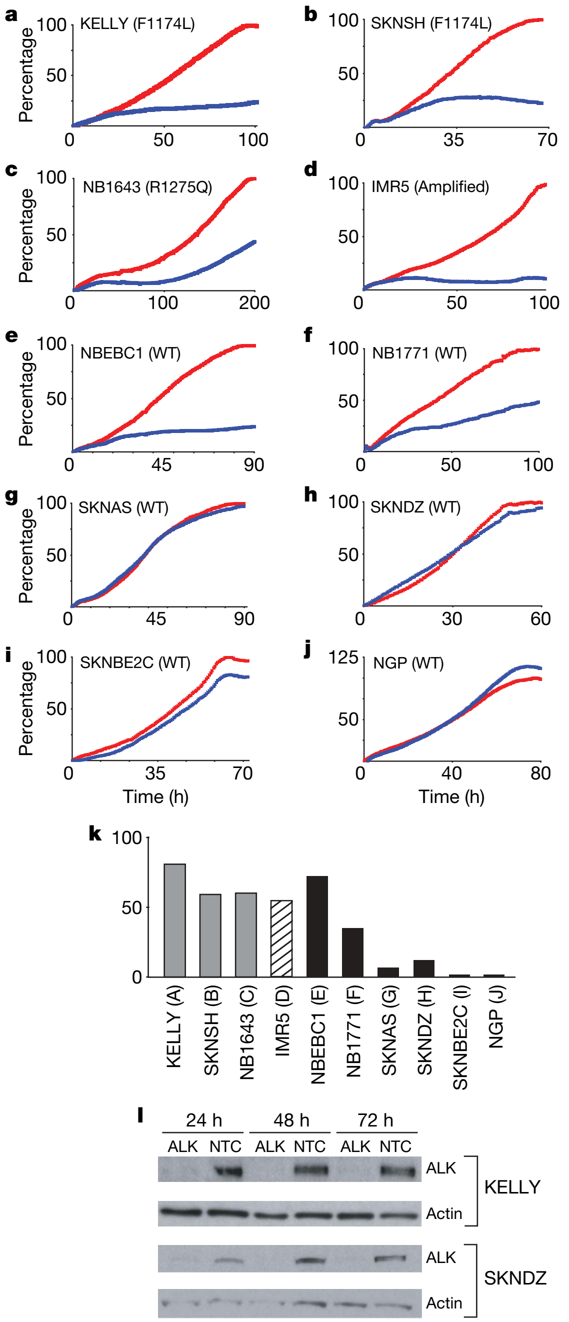


Figure 5 | *ALK* knockdown results in growth inhibition of *ALK* mutated or amplified neuroblastoma cell lines. a–j, Cellular growth for ten neuroblastoma cell lines that were transfected with siRNAs against *ALK* (blue) or *GAPDH* (red; two negative controls and one positive control not shown for clarity). The *x* axis is time in hours after transfection; the *y* axis is the percentage growth normalized to the siRNA against *GAPDH*. **k,** Summary of the percentage growth inhibition with *ALK* siRNA knockdown by *ALK* mutational and allelic status. **l,** Immunoblot showing a time course of ALK protein knockdown in the cell lines KELLY and SKNDZ. NTC, non-targeting control.

METHODS SUMMARY

Twenty probands with neuroblastoma and a family history of the disease were identified. Eight pedigrees had three or more affected individuals; six pedigrees contained only two affected individuals, but of first degree relation; and six pedigrees consisted of only two affected individuals, but of second, third or ≥fourth degree relationship. A total of 176 individuals (49 affected with neuroblastoma) were genotyped genome-wide, and two families were excluded owing to insufficient DNA for genotyping. We simulated marker data under a model of genetic homogeneity and autosomal dominant inheritance, and analysed the data using an affected-only approach comparable to the model-free approach used in the actual linkage analysis. Genotype data were checked for Mendelian inconsistencies using PedStats[45], and analysed for linkage using Merlin[46] and Lamp[47]. Regional candidates were re-sequenced using Sanger methodology. Predictions on the probability that DNA sequence alterations encode a mutant protein were performed using a support vector machine-based statistical classifier[26,27]. Four-hundred-and-ninety-one primary tumour samples and 27 cell lines were used for whole genome SNP-array analyses (550K) to determine copy number alterations[37]. mRNA knockdown of *ALK* and control targets was achieved with siRNAs against each target. The effects of siRNA knockdown on substrate adherent growth was quantified using the RT-CES

microelectronic cell sensor system (ACEA)[48,49]. Whole cell lysates were collected from the cell lines and from cells treated with either *ALK* siRNA or non-targeting control after transfection. Proteins were separated by SDS–PAGE gels and immunoblotted using anti-ALK and anti-phospho-ALK antibodies.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 April; accepted 14 July 2008.
Published online 24 August 2008; corrected 16 October 2008 (details online).

1. Maris, J. M., Hogarty, M. D., Bagatell, R. & Cohn, S. L. Neuroblastoma. *Lancet* **369,** 2106–2120 (2007).
2. Matthay, K. K. *et al.* Treatment of high-risk neuroblastoma with intensive chemotherapy, radiotherapy, autologous bone marrow transplantation, and 13-cis-retinoic acid. Children's Cancer Group. *N. Engl. J. Med.* **341,** 1165–1173 (1999).
3. Schwab, M. *et al.* Chromosome localization in normal human cells and neuroblastomas of a gene related to c-*myc*. *Nature* **308,** 288–291 (1984).
4. Attiyeh, E. F. *et al.* Chromosome 1p and 11q deletions and outcome in neuroblastoma. *N. Engl. J. Med.* **353,** 2243–2253 (2005).
5. Knudson, A. G. J. & Strong, L. C. Mutation and cancer: Neuroblastoma and pheochromocytoma. *Am. J. Hum. Genet.* **24,** 514–522 (1972).
6. Kushner, B. H., Gilbert, F. & Helson, L. Familial neuroblastoma. Case reports, literature review, and etiologic considerations. *Cancer* **57,** 1887–1893 (1986).
7. Maris, J. M. *et al.* Molecular genetic analysis of familial neuroblastoma. *Eur. J. Cancer* **33,** 1923–1928 (1997).
8. Friedman, D. L. *et al.* Increased risk of cancer among siblings of long-term childhood cancer survivors: a report from the childhood cancer survivor study. *Cancer Epidemiol. Biomarkers Prev.* **14,** 1922–1927 (2005).
9. Maris, J. M. & Brodeur, G. M. in *Neuroblastoma* (eds Cheung, N.-K. V. & Cohn, S. L.) 21–26 (Springer, 2005).
10. Longo, L. *et al.* Genetic predisposition to familial neuroblastoma: identification of two novel genomic regions at 2p and 12p. *Hum. Hered.* **63,** 205–211 (2007).
11. Maris, J. M. *et al.* Evidence for a hereditary neuroblastoma predisposition locus at chromosome 16p12–13. *Cancer Res.* **62,** 6651–6658 (2002).
12. Perri, P. *et al.* Weak linkage at 4p16 to predisposition for human neuroblastoma. *Oncogene* **21,** 8356–8360 (2002).
13. Amiel, J. *et al.* Polyalanine expansion and frameshift mutations of the paired-like homeobox gene *PHOX2B* in congenital central hypoventilation syndrome. *Nature Genet.* **33,** 459–461 (2003).
14. Mosse, Y. P. *et al.* Germline *PHOX2B* mutation in hereditary neuroblastoma. *Am. J. Hum. Genet.* **75,** 727–730 (2004).
15. Trochet, D. *et al.* Germline mutations of the paired-like homeobox 2B (*PHOX2B*) gene in neuroblastoma. *Am. J. Hum. Genet.* **74,** 761–764 (2004).
16. Raabe, E. H. *et al.* Prevalence and functional consequence of *PHOX2B* mutations in neuroblastoma. *Oncogene* **27,** 469–476 (2008).
17. van Limpt, V. *et al.* The Phox2B homeobox gene is mutated in sporadic neuroblastomas. *Oncogene* **23,** 9280–9288 (2004).
18. Weiss, W. A., Aldape, K., Mohapatra, G., Feuerstein, B. G. & Bishop, J. M. Targeted expression of *MYCN* causes neuroblastoma in transgenic mice. *EMBO J.* **16,** 2985–2995 (1997).
19. Osajima-Hakomori, Y. *et al.* Biological role of anaplastic lymphoma kinase in neuroblastoma. *Am. J. Pathol.* **167,** 213–222 (2005).
20. George, R. E. *et al.* Genome-wide analysis of neuroblastomas using high-density single nucleotide polymorphism arrays. *PLoS ONE* **2,** e255 (2007).
21. Morris, S. W. *et al.* Fusion of a kinase gene, *ALK*, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science* **263,** 1281–1284 (1994).
22. Griffin, C. A. *et al.* Recurrent involvement of 2p23 in inflammatory myofibroblastic tumors. *Cancer Res.* **59,** 2776–2780 (1999).
23. Jazii, F. R. *et al.* Identification of squamous cell carcinoma associated proteins by proteomics and loss of beta tropomyosin expression in esophageal cancer. *World J. Gastroenterol.* **12,** 7104–7112 (2006).
24. Soda, M. *et al.* Identification of the transforming *EML4–ALK* fusion gene in non-small-cell lung cancer. *Nature* **448,** 561–566 (2007).
25. Rikova, K. *et al.* Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* **131,** 1190–1203 (2007).
26. Torkamani, A. & Schork, N. J. Accurate prediction of deleterious protein kinase polymorphisms. *Bioinformatics* **23,** 2918–2925 (2007).
27. Torkamani, A. & Schork, N. J. Prediction of cancer driver mutations in protein kinases. *Cancer Res.* **68,** 1675–1682 (2008).
28. Ikenoue, T. *et al.* Functional analysis of mutations within the kinase activation segment of *B-Raf* in human colorectal tumors. *Cancer Res.* **63,** 8132–8137 (2003).
29. Ikenoue, T. *et al.* Different effects of point mutations within the *B-Raf* glycine-rich loop in colorectal tumors on mitogen-activated protein/extracellular signal-regulated kinase kinase/extracellular signal-regulated kinase and nuclear factor κB pathway and cellular transformation. *Cancer Res.* **64,** 3428–3435 (2004).
30. Kannan, N. & Neuwald, A. F. Did protein kinase regulatory mechanisms evolve through elaboration of a simple structural component? *J. Mol. Biol.* **351,** 956–972 (2005).
31. Jeffers, M. *et al.* Activating mutations for the met tyrosine kinase receptor in human cancer. *Proc. Natl Acad. Sci. USA* **94,** 11445–11450 (1997).
32. Balak, M. N. *et al.* Novel D761Y and common secondary T790M mutations in epidermal growth factor receptor-mutant lung adenocarcinomas with acquired resistance to kinase inhibitors. *Clin. Cancer Res.* **12,** 6494–6501 (2006).
33. Lee, J. W. *et al.* *ERBB2* kinase domain mutation in the lung squamous cell carcinoma. *Cancer Lett.* **237,** 89–94 (2006).
34. Lee, J. W. *et al.* Somatic mutations of *ERBB2* kinase domain in gastric, colorectal, and breast carcinomas. *Clin. Cancer Res.* **12,** 57–61 (2006).
35. Wang, Q. *et al.* Integrative genomics identifies distinct molecular classes of neuroblastoma and shows that multiple genes are targeted by regional alterations in DNA copy number. *Cancer Res.* **66,** 6050–6062 (2006).
36. Vogelstein, B. & Kinzler, K. W. Cancer genes and the pathways they control. *Nature Med.* **10,** 789–799 (2004).
37. Maris, J. M. *et al.* Chromosome 6p22 locus associated with clinically aggressive neuroblastoma. *N. Engl. J. Med.* **358,** 2585–2593 (2008).
38. Iwahara, T. *et al.* Molecular characterization of ALK, a receptor tyrosine kinase expressed specifically in the nervous system. *Oncogene* **14,** 439–449 (1997).
39. Chiarle, R., Voena, C., Ambrogio, C., Piva, R. & Inghirami, G. The anaplastic lymphoma kinase in the pathogenesis of cancer. *Nature Rev. Cancer* **8,** 11–23 (2008).
40. Lamant, L. *et al.* Expression of the *ALK* tyrosine kinase gene in neuroblastoma. *Am. J. Pathol.* **156,** 1711–1721 (2000).
41. Motegi, A., Fujimoto, J., Kotani, M., Sakuraba, H. & Yamamoto, T. ALK receptor tyrosine kinase promotes cell growth and neurite outgrowth. *J. Cell Sci.* **117,** 3319–3329 (2004).
42. Miyake, I. *et al.* Activation of anaplastic lymphoma kinase is responsible for hyperphosphorylation of ShcC in neuroblastoma cell lines. *Oncogene* **21,** 5823–5834 (2002).
43. Dirks, W. G. *et al.* Expression and functional analysis of the anaplastic lymphoma kinase (*ALK*) gene in tumor cell lines. *Int. J. Cancer* **100,** 49–56 (2002).
44. McDermott, U. *et al.* Genomic alterations of anaplastic lymphoma kinase may sensitize tumors to anaplastic lymphoma kinase inhibitors. *Cancer Res.* **68,** 3389–3395 (2008).
45. Wigginton, J. E. & Abecasis, G. R. PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics* **21,** 3445–3447 (2005).
46. Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. Merlin–rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genet.* **30,** 97–101 (2002).
47. Li, M., Boehnke, M. & Abecasis, G. R. Joint modeling of linkage and association: identifying SNPs responsible for a linkage signal. *Am. J. Hum. Genet.* **76,** 934–949 (2005).
48. Yu, N. *et al.* Real-time monitoring of morphological changes in living cells by electronic cell sensor arrays: an approach to study G protein-coupled receptors. *Anal. Chem.* **78,** 35–43 (2006).
49. Cole, K. A. *et al.* A functional screen identifies miR-34a as a candidate neuroblastoma tumor suppressor gene. *Mol. Cancer Res.* **6,** 735–742 (2008).
50. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods* **25,** 402–408 (2001).

Acknowledgements We acknowledge the families and children that participated in this research study, and the Children's Oncology Group for providing specimens. We thank W. London for providing statistical analyses related to the Children's Oncology Group tumour set, H. Rydbeck for his assistance with the linkage analysis, M. Wasik for technical assistance, and J. Felgenhauer, N. Van Roy and C. McConville for providing neuroblastoma pedigrees. This work was supported in part by National Institutes of Health grants K08-111733 (Y.P.M.), R01-CA78454 (J.M.M.), R01-CA87847 (J.M.M.), an American Society of Clinical Oncology Career Development Award (Y.P.M.), the Foerderer-Murray Fund (Y.P.M.), the Carly Hillman Fund (Y.P.M.), the Alex's Lemonade Stand Foundation (J.M.M.), the Andrew's Army Foundation (J.M.M.), the Giulio D'Angio Endowed Chair (J.M.M.), the Italian Neuroblastoma Foundation (L.L.), the Center for Applied Genomics at the Joseph Stokes Research Institute (H.H.), Scripps Genomic Medicine (A.T., N.J.S.), the Scripps Dickinson Scholarship (A.T.), and the Abramson Family Cancer Research Institute (J.M.M.).

Author Contributions Y.P.M. and J.M.M. designed the experiments and wrote the manuscript. Y.P.M., M.L., J.M.M., G.L., F.S., P.P. and G.P.T. collected the families for the linkage analysis. Y.P.M., M.L., L.L., C.K., C.H., E.R., H.H. and M.D. performed the genome-wide genotyping and linkage analysis. M.L. performed the DNA sequencing and analyses. K.A.C., A.W. and M.J.L. performed the siRNA experiments. E.F.A., H.H. and Y.P.M. performed the tumour SNP genotyping/copy number analyses. J.E.L., K.A.C. and A.W. performed the expression analyses. K.A.C., R.S. and M.L. performed the protein work. G.M.B. initiated the collection of neuroblastoma pedigrees. A.T. and N.J.S. performed the structural analysis of *ALK* coding mutations.

Author Information All sequence variations have been deposited to GenBank under accession numbers EU660517 to EU660527. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.M.M. (maris@chop.edu).

METHODS

Research subjects and samples. Families with a history of neuroblastoma in at least one other relative were eligible to participate. Only germline DNA from the neuroblastoma pedigrees was studied as no tumour tissue was available. Sporadic neuroblastoma tumour samples with matched constitutional DNA were acquired from the Children's Oncology Group neuroblastoma tumour bank. The Children's Hospital of Philadelphia Institutional Review Board approved this research.

Linkage analysis. A genome-wide linkage scan was performed using the Illumina Linkage IVb SNP panel. Genotype data were checked for Mendelian inconsistencies using PedStats[45], and analysed for linkage using Merlin[46] and Lamp[47]. The genome-wide screen for linkage was performed with both maximum-likelihood allele frequency estimates and model-free analyses. Because the pattern of inheritance is complex, we used a model-free approach so as not to assume any specific mode of inheritance. Model-based analyses were performed for all SNPs included in the critical region under a dominant mode of inheritance, assuming four different gene frequencies (0.0001, 0.001, 0.01 and 0.1) and dominant transmission with varying penetrance of the disease across a broad range, from 0.0001 to 0.68. Model-based linkage analysis was also performed using the method implemented in LAMP, assuming a prevalence of the disease of 0.000143 (1 in 7,000), and maximizing the lod scores over all possible disease models (MOD score analysis). In every analysis, the critical interval was defined as the region with associated lod scores that were greater than the maximum lod score minus three. Because linkage disequilibrium among markers is known to inflate the lod scores from linkage analysis in the presence of missing founders, its impact on the lod scores was assessed at the chromosome 2p critical interval. To model linkage disequilibrium, we organized markers into clusters by means of Merlin, which uses population haplotype frequencies derived from the HapMap project (http://www.hapmap.org/).

DNA sequencing. Shotgun resequencing from templates generated by long PCR for an 18 kb region surrounding the *MYCN* locus was performed using a 454 Life Sciences instrument (Branford) after bi-directional sequencing of the three coding exons showed no disease causal sequence variations in the pedigrees. Bi-directional sequencing of the *ALK* coding sequence (primer sequences available on request) was performed in the following distinct sample sets: (1) constitutional DNA from the proband and an unaffected first degree relative from the 20 neuroblastoma pedigrees, with repeat sequencing of amplicons containing any DNA sequence variations, and sequencing of amplicons containing confirmed variations in the remaining family members; (2) 27 human neuroblastoma-derived cell line DNAs maintained at the Children's Hospital of Philadelphia (available on request); (3) tumour DNA from 167 sporadic neuroblastomas from the Children's Oncology Group tumour bank; and (4) 109 normal constitutional DNAs from the SNP500 Cancer Resource panel purchased from the Coriell Institute for Medical Research. To verify neuroblastoma cell line integrity, all lines were routinely genotyped (AmpFLSTR Identifiler kit; Applied Biosystems) and tested for mycoplasma.

Mutation prediction. Cancer mutant predictions and analysis were performed as described previously[27]. In brief, a support vector machine was trained on common SNPs (presumed neutral) and congenital disease-causing SNPs characterized by a variety of sequence, structural and phylogenetic parameters. Training and predictions were performed using somatic mutations occurring in and outside the kinase catalytic core separately. The support vector machine-based method was then applied to the *ALK* mutants, and the probability that each mutant is a driver was computed using the support vector machine. The threshold taken for calling a SNP a driver was taken to be 0.49 for catalytic domain mutations and 0.53 for all other mutations[26]. For comparison to previously observed cancer mutations, *ALK* mutants were mapped to positions of a catalytic core alignment generated with characteristic site motifs, and previously observed cancer mutants mapping to the same positions were noted[27].

Tumour copy number analysis. Tumour samples were assayed on the Illumina Infinium II HumanHap550 BeadChip technology, at the Center for Applied Genomics at the Children's Hospital of Philadelphia. A total of 750 ng of genomic DNA was used as the input in each case; the assay was performed and data were analysed following the manufacturer's recommendations and as previously described[37].

Quantitative mRNA expression. Relative *ALK* expression was determined using the $2^{-\Delta\Delta C_T}$ method[50], using *GAPDH* as the endogenous control and using the second dCT (change in cycle threshold values) as the lowest expressed cell line, using methods as previously described[49].

***ALK* siRNA knockdown in neuroblastoma cell lines.** A total of $1–5 \times 10^4$ neuroblastoma cells were plated in triplicate overnight in antibiotic-free complete media in the 96-well RT-CES microelectronic cell sensor system (ACEA)[48,49]. Cells were transiently transfected with 200 µl containing 50 nM of pooled siRNAs (four separate siRNAs per transcript targeted) against *ALK* (M-003103-02), *GAPDH* (D-001140-01-20) negative control, non-targeting negative control, or *PLK1* (M-003290-01) positive control (siGENOME SMARTpool siRNA, Dharmacon) using 0.1% (v/v) Dharmafect I, according to the manufacturer's protocol (Dharmacon). The four separate *ALK*-directed siRNA sense-direction sequences are: GGGCCUGUAUACCGGAUAAUU (ALK J-003103-10), GUGCCAUGCUGCCAGUUAAUU (ALK J-003103-11), CCGCUUUGCCGAUAGAAUAUU (ALK J-003103-12) and GGAGCCACCUACGUAUUUAUU (ALK J-003103-13). In brief, 35 µl of 1 µM siRNA and 35 µl of serum-free media were combined with 0.7 µl Dharmafect I in 70 µl of serum-free media, incubated for 20 min at room temperature, and then 560 µl of antibiotic-free complete media was added. The culture media was gently removed from the plated cells and replaced by 200 µl of fresh media containing the siRNA, mock or complete media. Cell growth was monitored continuously and recorded as a cell index every 30 min for a minimum of 96 h. The cell index is derived from the change in electrical impedance as the living cells interact with the biocompatible microelectrode surface in the microplate well effectively measuring cell number, shape and adherence. Forty-eight hours after siRNA transfection, total RNA was extracted from the cells that had been plated in a parallel 96-well plate using the Qiagen mini extraction kit, with DNAase treatment. Two hundred nanograms of total RNA was oligodT primed and reverse transcribed using Superscript II reverse transcriptase (Invitrogen). *ALK*, *HPRT*, *GAPDH* and *PPIB* expression levels were measured by quantitative RT–PCR using Taqman gene expression assays (ABI), quantified on corresponding standard curves and normalized to the geometric mean of the three housekeeping genes. Two independent experiments were performed in triplicate. Growth inhibition of the neuroblastoma cell lines was determined by comparing the growth curve for *ALK* siRNA to the curve for *GAPDH* siRNA at the time of maximum cell index (CI_{max}): % growth inhibition = $(1 - CI_{ALK\ siRNA} / CI_{GAPDH\ siRNA}) \times 100$. *ALK* and *GAPDH* knockdown were determined by comparing the relative *ALK* expression: % knockdown = $(1 - ALK_{ALK\ siRNA}/ALK_{Control\ siRNA}) \times 100$. The average percentage knockdown of *ALK* across all cell lines was 60% (range 21%–86%). The average percentage knockdown of *GAPDH* was 75% (range 61%–95%).

ALK protein and phosphoprotein detection. Neuroblastoma cell lines were grown in T75 flasks under standard cell culturing conditions. For KELLY and SKNDZ lysates collected from the siRNA knockdown experiments, cells were plated in T25 flasks, transfected with 10 nM siRNA (as before) and collected at 24, 48 and 72 h after transfection. At 60%–80% confluency (or the appropriate time point), the cells were collected, pelleted and washed twice with ice cold PBS. Whole cell lysates were extracted with 100 µl cell extraction buffer (FNN011, Invitrogen) containing protease inhibitors (P-2714, Sigma) and phenylmethyl sulphonyl fluoride, briefly sonicated and rotated for 1 h at 4 °C. After a 30 min centrifugation at 4 °C, the supernatant was removed and protein quantification was performed using the Bradford method. Lysates (50 µg for siRNA experiment and 100 µg for native cell lines) were separated on 4%–12% Bis–Tris gradient gels and transferred to PVDF membranes. Membranes were then washed and incubated with anti-ALK (1:1,000; Cell Signaling, 3333), anti-phospho-ALK (1:1,000; Cell Signaling, 3341) and anti-actin (1:5,000; Santa Cruz, sc-2352) antibodies according to the cell signalling western protocol.

ARTICLES

Structure of a complex of the ATPase SecA and the protein-translocation channel

Jochen Zimmer[1], Yunsun Nam[1] & Tom A. Rapoport[1]

Most proteins are secreted from bacteria by the interaction of the cytoplasmic SecA ATPase with a membrane channel, formed by the heterotrimeric SecY complex. Here we report the crystal structure of SecA bound to the SecY complex, with a maximum resolution of 4.5 ångström (Å), obtained for components from *Thermotoga maritima*. One copy of SecA in an intermediate state of ATP hydrolysis is bound to one molecule of the SecY complex. Both partners undergo important conformational changes on interaction. The polypeptide-cross-linking domain of SecA makes a large conformational change that could capture the translocation substrate in a 'clamp'. Polypeptide movement through the SecY channel could be achieved by the motion of a 'two-helix finger' of SecA inside the cytoplasmic funnel of SecY, and by the coordinated tightening and widening of SecA's clamp above the SecY pore. SecA binding generates a 'window' at the lateral gate of the SecY channel and it displaces the plug domain, preparing the channel for signal sequence binding and channel opening.

Many bacterial proteins are transported across or are integrated into the plasma membrane, a process that is similar to protein translocation across the endoplasmic reticulum (ER) membrane in eukaryotes (for review, see ref. 1). Translocation begins with a targeting phase, during which proteins are directed to the membrane by hydrophobic sequences, which are either cleavable signal sequences or transmembrane (TM) segments of membrane proteins. Subsequently, the soluble proteins (such as secretory proteins) are transported through a hydrophilic channel across the membrane, whereas the hydrophobic TM segments of membrane proteins exit sideways through a lateral gate of the channel into the surrounding lipid phase. The channel is formed from a conserved heterotrimeric membrane protein complex, known as the SecY complex in bacteria and archaea, and the Sec61 complex in eukaryotes[1]. The complex consists of a large α-subunit (SecY or Sec61p), and two smaller β- and γ-subunits (called SecG and SecE in bacteria). The channel can associate with different partners that provide a driving force for translocation. In bacteria, the SecY channel can either bind to the ribosome to translocate polypeptides during their synthesis (co-translational translocation), or associate with the cytoplasmic SecA ATPase to transport secretory proteins after completion of their synthesis (post-translational translocation). SecA is also required for the biosynthesis of some membrane proteins[2–4]. How the channel connects to its different partners, and how polypeptides are moved through the channel, is unclear.

The crystal structure of an archaeal SecY complex (from *Methanococcus jannaschii*) and previous biochemical experiments have shown that a single SecY copy forms the active channel[5–7]. SecY consists of two linked halves, TM1–TM5 and TM6–TM10, which form a lateral gate at the front and are clamped together at the back by SecE[5]. The helices of SecY form an hourglass-shaped pore that consists of cytoplasmic and external funnels. The constriction of the pore is located about halfway across the membrane and is formed by a 'pore ring' of hydrophobic amino acids that project their side chains radially inwards. The cytoplasmic funnel is empty, whereas the external funnel is plugged by a short helix that interacts with the pore residues. This crystal structure therefore represents a closed state of SecY. Opening of the channel involves the displacement of the plug[8,9], which is triggered by the intercalation of a hydrophobic sequence (a signal or a TM sequence) into the lateral gate of the channel[10].

Post-translational translocation in bacteria requires only the SecY complex and the ATPase SecA, which 'pushes' polypeptides through the SecY channel[11–13]. SecA consists of two RecA-like nucleotide-binding domains (NBD1 and NBD2), which bind the nucleotide between them, a polypeptide-cross-linking domain (PPXD), a helical scaffold domain (HSD) and a helical wing domain (HWD)[14]. Although several crystal structures of isolated SecA have been determined, the function of the different domains and the mechanism by which SecA moves polypeptides through the channel remain unknown. Disulphide cross-linking experiments suggest that SecA binds by its NBD1 domain to a non-translocating SecY copy, and moves the polypeptide chain through a neighbouring SecY molecule[6]. These and other experiments indicate that SecA functions as a monomer during translocation[7,15–17], but the issue remains controversial[18–20].

Here we report crystal structures of SecA bound in an intermediate state of nucleotide hydrolysis to the SecY channel. The structures suggest mechanisms for how the channel is opened and prepared for the arrival of a translocation substrate, and how SecA moves polypeptides through the channel.

Structure determination of SecA–SecY complexes

We crystallized complexes containing *Bacillus subtilis* SecA without its non-essential carboxy-terminal domain, and either *Thermotoga maritima* SecYE or *Aquifex aeolicus* SecYEG. These crystals diffracted X-rays to a maximum resolution of 6.2 Å and 7.5 Å, respectively. A higher resolution data set (4.5 Å) was obtained for a complex in which both partners were from *T. maritima* and the SecYEG complex was seleno-methionine (Se-Met) derivatized. All complexes were crystallized in the detergent Cymal-6 in the presence of ADP and BeF_x. The structure of the complex of *B. subtilis* SecA and *T. maritima* SecYE was determined by molecular replacement with a *B. subtilis* SecA structure[21] and served as an initial model for the other complexes. The building of a 4.5 Å resolution model of the *T. maritima* SecA–SecY complex was facilitated by the Se-Met positions (Supplementary Fig. 1), and by the high quality of the phases, leading to an electron density map that allowed the identification of large

[1]Howard Hughes Medical Institute and Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA.

amino acid side chains (Fig. 1a and Supplementary Fig. 2). Model building also took into account conserved interactions between amino acids in previously determined SecA and SecY structures[5,21,22] (sequence alignments are shown in Supplementary Figs 3 and 4). The final structure was refined to R_{work} and R_{free} factors of 27.9% and 30.3% (Table 1), respectively, and was used for all interpretations. It comprises all residues of SecA and most residues of SecYEG. No model could be built for the periplasmic loop between TM1 and TM2a of SecY (residues 42–61), as well as for residues of some termini (SecY residues 1–7 and 424–431; SecE residues 1–9; SecG residues 1–8 and 74–76). Furthermore, there are uncertainties about the tip of the loop between TM6 and TM7 (residues 240–254). An ADP–BeF_3^- complex was modelled into the electron density observed in the nucleotide-binding pocket of SecA (Supplementary Fig. 5).

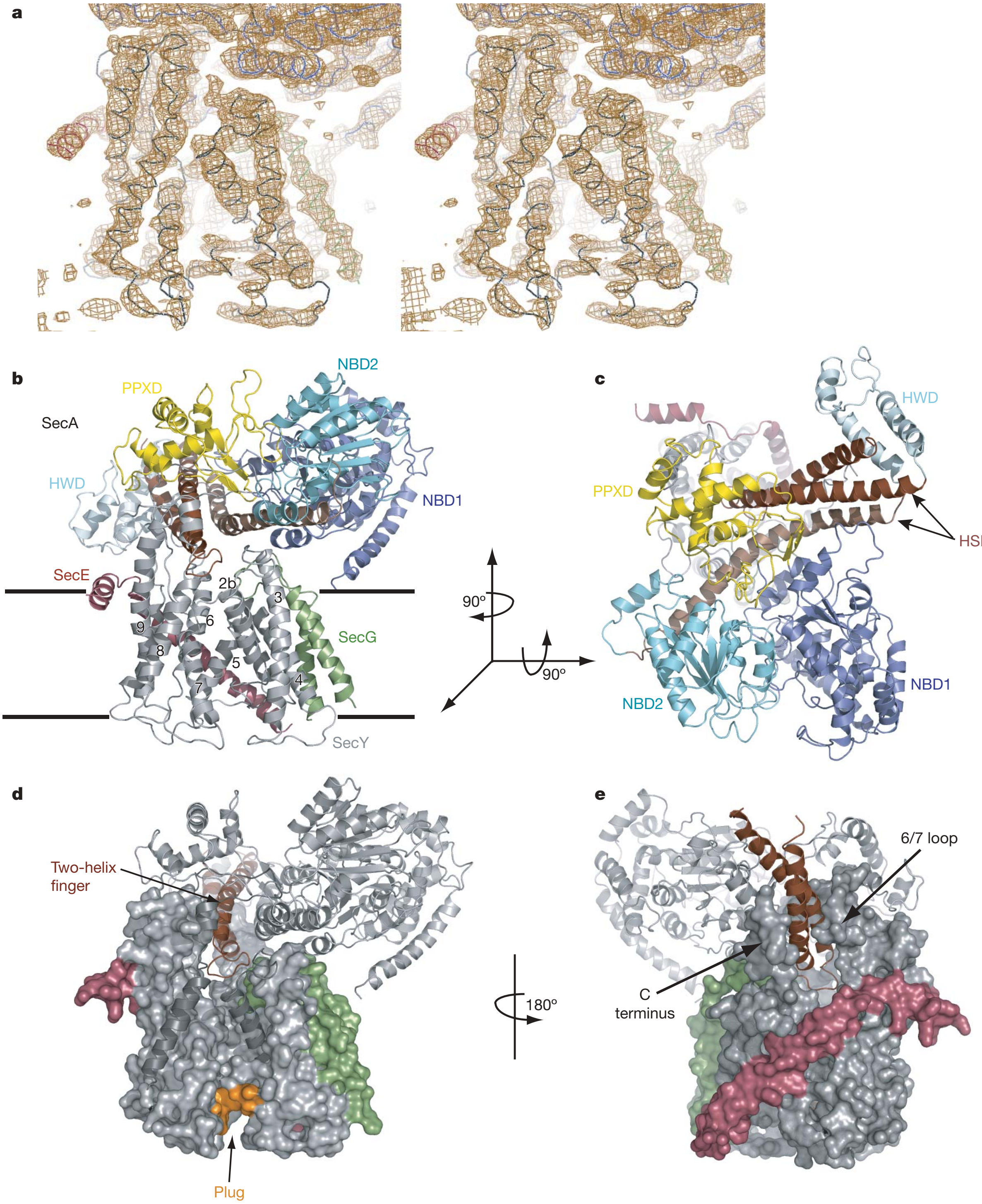


Figure 1 | Architecture of the *T. maritima* SecA–SecYEG complex. a, Stereo view of a σA-weighted, phase combined, NCS averaged, and B-factor sharpened $2F_o - F_c$ electron density map (contoured at 1σ). The view of the lateral gate of SecY is shown, with the Cα-trace of SecY in grey, SecE in red, SecG in green and SecA in blue. **b**, Cartoon of the complex viewed from the side. The lines indicate the membrane boundaries. **c**, As in **b**, but viewed from the cytoplasm. **d**, The two-helix finger of SecA inside the cytoplasmic funnel of SecY. TM2b and TM8, as well as the tip of the 6–7 loop, are shown as cartoons for clarity. Plug residues are coloured in orange. **e**, As in **d**, but shown from the back.

Architecture of the SecA–SecYEG complex

The structures obtained with SecA and SecY from three different species all have the same architecture (Fig. 1b, c and Supplementary Figs 6 and 7), despite forming different crystal contacts. Crystals were only obtained in the presence of ADP and either BeF_x or AlF_x, corresponding to intermediate states during ATP hydrolysis. Gel filtration experiments indicated that the interaction between SecA and SecY is strongest under these conditions (Supplementary Fig. 8).

The structures show one SecA molecule bound to one copy of the SecY channel (Fig. 1b, c). The flat SecA molecule is oriented approximately parallel to the plane of the membrane (Fig. 1b). Viewed from the cytosol, NBD1, NBD2, PPXD and HSD sit at the corners of an approximate square (Fig. 1c). NBD1 is linked diagonally with the PPXD. Similarly, the linked NBD2 and HSD are located in opposite corners of the square.

The extensive interactions of SecA with the SecY channel cover approximately 6,800 $Å^2$ of surface area. The most crucial interactions occur between the PPXD of SecA and the 8–9 and 6–7 loops of SecY. Mutations in these loops indeed affect SecA-mediated translocation[23,24], and SecA constructs lacking the PPXD fail to form stable complexes with SecY (data not shown). The HSD of SecA also makes extensive contacts with SecY, accounting for almost 50% of the buried surface area. The interaction of the C-terminal tail of SecY with the HSD is probably functionally important, as indicated by the detrimental effect of SecY truncations and point mutations[25,26] and by the interference of a C-terminal His-tag on SecY with SecA interaction (Y.N., unpublished observations). The long helix of the HSD rests on the 2b–3 loop of SecY and the SecG loop. Both interactions are probably not strong as the contact surfaces are small, the 2b–3 loop contains only a few conserved amino acids, and SecG is not essential for translocation. Minor interactions also exist between SecG and the NBD1 of SecA (residues 420–430 and the amino terminus). NBD2 makes no contacts with the SecY channel.

A two-helix finger, which corresponds to the shorter helices of the HSD of SecA, is inserted into the cytoplasmic funnel of the channel (Fig. 1d, e). The loop connecting the two helices is located right above the entrance to the SecY pore. The finger has an angle of $\sim 45°$ with respect to the plane of the membrane and rests in a cleft between the 6–7 loop and the C-terminal tail of SecY (Fig. 1e).

The structure is consistent with our previous proposal that SecA functions as a monomer during translocation[15]. The single SecY copy seen in the structure must be the one through which polypeptide chains are moved during translocation. Disulphide-bridge cross-linking shows that the 6–7 loop of a non-translocating SecY copy contacts a face of NBD1 that is exposed in our structure (residues 41 and 48 in *T. maritima* SecA)[6]. This copy is lost on detergent solubilization[7]. It

Table 1 | Crystallographic statistics

Data set	*Tm* SecA–SecYEG	*Tm* SecA–SecYEG	*Bs* SecA–*Tm* SecYE	*Bs* SecA–*Aa* SecYEG
Data collection				
Compound	Se-Met SecYEG	Native	Native	Native
Resolution limit (Å)	4.5	5.7	6.2	7.5
Space group	$P2_12_12_1$	$P2_1$	$P6_5$	$P2_12_12_1$
Cell constants (Å)/(°)	101.6, 156.0, 358.1	172.5, 102.5, 181.9	240.3, 240.3, 179.5	146.4, 168.0, 187.7
(a, b, c, / α, β, γ)	90, 90, 90	90, 100.9, 90	90, 90, 120	90, 90, 90
Measured reflections	76,164	39,014	28,484	6,238
Unique reflections	34,733	18,704	13,510	5,416
Mean $I/\sigma(I)$	16.1 (2.0)*	16.5 (1.5)	26.2 (2.5)	16.2 (1.6)
R_{sym} (%)†	0.065 (0.85)	0.055 (0.87)	0.075 (>1.0)	0.072 (0.76)
Completeness (%)	97.7 (89.7)	93.3 (81.8)	100.0 (100.0)	99.1 (99.8)
Redundancy	11.6 (8.8)	3.8 (2.5)	9.4 (9.0)	3.5 (3.6)
SAD-phasing‡				
FOM (from SOLVE)	0.25 (0.12)			
FOM (post NCS average)	0.72 (0.27)			
Refinement			Rigid body	Rigid body
Resolution range (Å)	15–4.5		15–6.2	15–7.5
NCS§	2	2	2	2
R_{work} (%)#	27.9		42.0	38.8
R_{free} (%)*	30.3		42.7	39.3
Number of reflections				
Total	32,721		9,176	4,884
R_{free}	3,269		1,012	532
Number of atoms‖				
Protein	21,312			
Non-protein	62			
Model geometry				
r.m.s.d. bond length (Å)	0.006			
r.m.s.d. bond angle (°)	1.3			
Average B-factor ($Å^2$)				
Main chain	354			
Side chain	323			
Ramachandran analysis¶ (for 1,333 residues) (%)				
Most favoured	68.5			
Allowed	26.1			
Generously allowed	4.3			
Disallowed	1.1			

Aa, *A. aeolicus*; *Bs*, *B. subtilis*; *Tm*, *T. maritima*; SAD, single wavelength anomalous dispersion.
* Values in parenthesis refer to the highest resolution shell.
† R_{sym} is defined as $\Sigma|I_i - <I>|/\Sigma|I|$, where I_i is the intensity of an individual reflection and $<I>$ is the average intensity of that reflection.
‡ From 11 selenium sites identified per SecA–SecYEG complex in the resolution range from 15 to 6 Å. NCS averaging included phase extension to 4.5 Å.
§ Number of SecA–SecY complexes per crystallographic asymmetric unit.
\# R_{work} is defined as $\Sigma|F_o| - |F_c|/\Sigma|F_o|$ where F_c is the calculated and F_o is the observed structure factor amplitude.
* R_{free} is calculated according to R_{work} for a randomly selected subset containing 10% of reflections.
‖ In the asymmetric unit.
¶ For one SecA–SecYEG complex performed in PROCHECK.

could bind to the translocating SecY copy in an approximate back-to-back orientation, although its exact location is difficult to predict. In a membrane, the interaction between the two SecY complexes might be stabilized by negatively charged lipids, which are known to be essential for translocation[27,28]. Although SecA on its own can bind to lipid bilayers[28], SecA does not seem to make strong contacts with lipids when associated with the SecY channel, as only its extreme N terminus comes close to the membrane surface.

Conformational changes in SecA

Overall, the architecture of *T. maritima* SecA closely resembles the structure of *B. subtilis* SecA with which it shares ~45% sequence identity (Fig. 2a–c). The most notable difference is the insertion of a mainly α-helical loop at the periphery of NBD1 (residues 154–204; Fig. 2c). Crystal structures of SecA alone from different organisms show the molecule in two different conformations, with the PPXD either packed against the HWD or rotated away from it (closed and open conformation, Fig. 2a, b)[14,21]. When SecA binds to the SecY complex, the PPXD rotates even further and makes contact with NBD2 (Fig. 2c). With respect to the closed conformation, the overall movement of the PPXD is a large rigid-body rotation by ~80° towards NBD2 and a tilt of ~10° towards the membrane, with the hinge formed by two short β-strands connecting the PPXD and NBD1. An evolutionarily conserved loop in the PPXD (residues 360–370) now contacts both NBD1 and NBD2 and comes close to the nucleotide-binding site (Fig. 2c). The interface between the PPXD and NBD2 is perfectly aligned with the lateral gate of SecY, generating a continuous seam on the surface of the SecA–SecY complex (Supplementary Fig. 9).

NBD1 and NBD2 are closer to each other than in all structures of isolated SecA (Fig. 2d), which correspond to nucleotide-free or ADP-bound states. The relative positions of NBD1 and NBD2 in our structure are characteristic of the ATP-bound or transition states of RecA-like ATPases[29]. With NBD1 as a reference, the transition from the ADP state to the ADP–BeF_x state would involve a ~15° rigid body rotation of NBD2 about an axis that runs parallel to the plane of the membrane and through the nucleotide-binding cleft (Fig. 2d). The similarity with related ATPases suggests that the movement of the essential Arg finger of NBD2 (that is, Arg 570) towards the γ-phosphate triggers ATP hydrolysis[15].

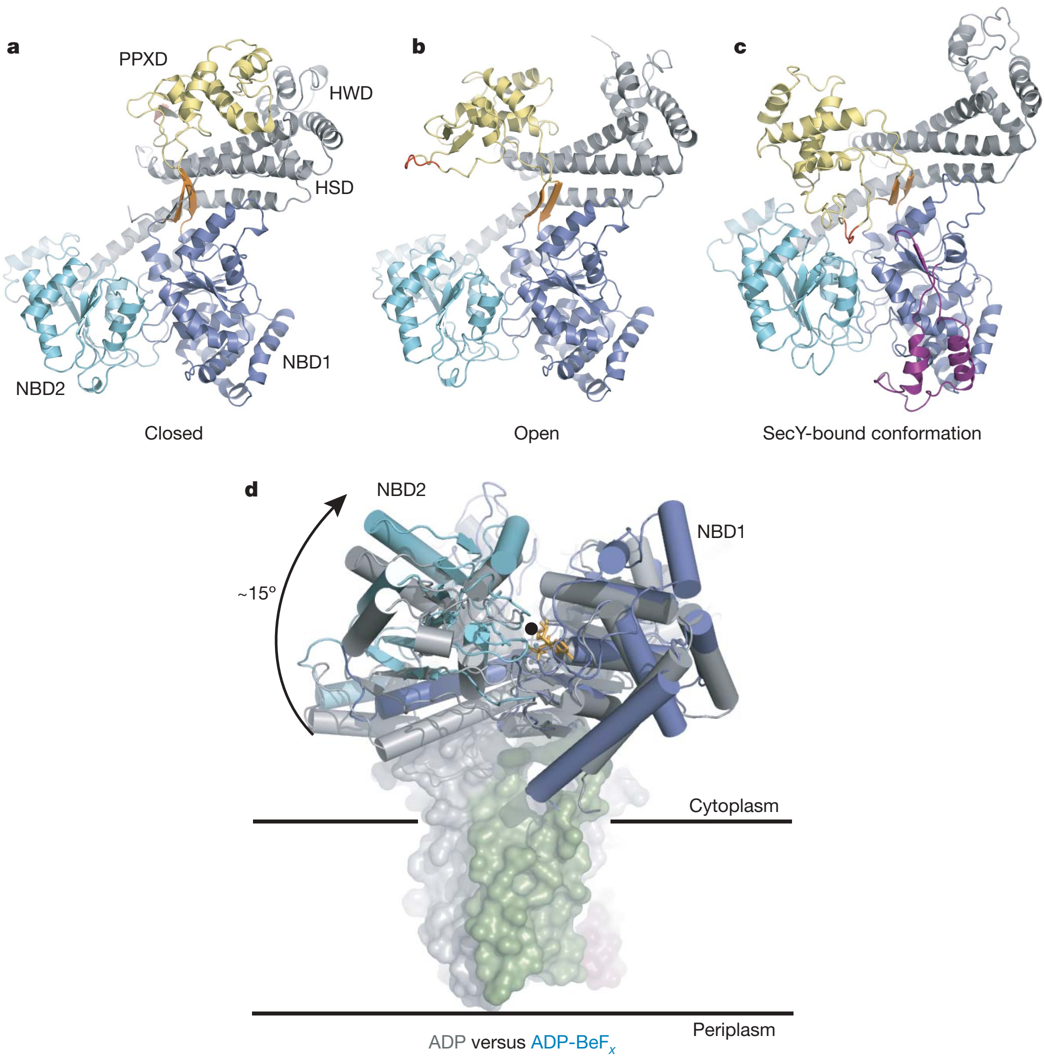


Figure 2 | Conformational changes of SecA. a–c, Different orientations of the PPXD relative to the NBDs. The view is from the cytosol, as in Fig. 1c. The closed (**a**) and open (**b**) conformations refer to the *B. subtilis* SecA structures (PDB accessions 1M6N and 1TF2, respectively). The loop of the PPXD contacting NBD1 and NBD2 in the SecA–SecY complex is highlighted in red. The insertion in the *T. maritima* NBD1 is shown in purple. The conserved β-strands connecting NBD1 with the PPXD are shown in orange. **d**, Position of the NBDs in the ADP-bound state of *B. subtilis* SecA (grey) versus the ADP–BeF_x bound state of *T. maritima* SecA (blue) when associated with SecY. The structures were aligned with respect to NBD1. The rotation axis indicated by a black circle is parallel to the plane of the membrane. The modelled ADP–BeF_3^- is shown as orange sticks.

Conformational changes in the SecY complex

The *T. maritima* SecY complex has a similar overall structure as the complex from *M. jannaschii*[5]. The main difference is the replacement of Secβ by SecG. SecG has two TMs with both termini located on the external side of the membrane (Fig. 1b), in agreement with previous topology determinations[30,31]. The N-terminal TM was assigned on the basis of Met 24, which is seen in an anomalous difference Fourier map of the Se-Met-derivatized SecY complex (Supplementary Fig. 1). The C-terminal TM has the same orientation and position as the single TM of Secβ in *M. jannaschii* (Fig. 3a). The SecG loop is sandwiched between SecA and SecY, explaining why the loop is protected against proteolytic cleavage during translocation[30].

Conformational changes in the SecY channel induced by SecA can best be appreciated by comparison of our structure with the crystal structure of the *M. jannaschii* SecY channel[5]. Although archaea lack SecA, the structure of the *M. jannaschii* SecY complex is remarkably similar to the one in SecA-containing organisms, represented by the ~8 Å structure of the *Escherichia coli* SecY complex[32,33]. Whereas the N-terminal half of *T. maritima* SecY aligns well with the corresponding region of *M. jannaschii* SecY, all helices in the C-terminal half shift outwards, with TM8 and TM9 showing the most notable changes (Fig. 3a). Without this conformational change, the two-helix finger of SecA would clash with the α-helical region of the 8–9 loop. The shift of TM9 causes the amphipathic helix of SecE, which lies on the cytoplasmic surface of the membrane, to also move outwards.

SecA binding generates important changes in the lateral gate of SecY that is formed from TM2b, TM3, TM7 and TM8. The displacement of TM7 opens a window towards the centre of the lipid bilayer (Fig. 3b). The gap between the side chains of TM7 and TM2b has a width of about 5 Å. On the cytoplasmic side of the membrane, the lateral gate seems to be closed towards the head groups of the phospholipids, as TM2b and TM8 remain in close proximity. The lateral gate is also closed towards the external side by the loop after TM2a.

The plug helix (TM2a) has moved away from the centre of the SecY channel towards the external side of the membrane (Fig. 3a and Supplementary Fig. 6). It no longer contacts the pore residues, and instead packs against the C terminus of TM7. Despite its displacement, the plug still closes the channel (Fig. 1d), probably hindering even the permeation of water and ions. The plug sits at the front of the channel and is not contacting the TM segment of SecE, as it does during translocation[8,9].

Implications for protein translocation

Translocation seems to start in the cytosol with an interaction of SecA with a signal-sequence-containing polypeptide chain, a process that is often mediated by the chaperone SecB (for review, see ref. 1). Next,

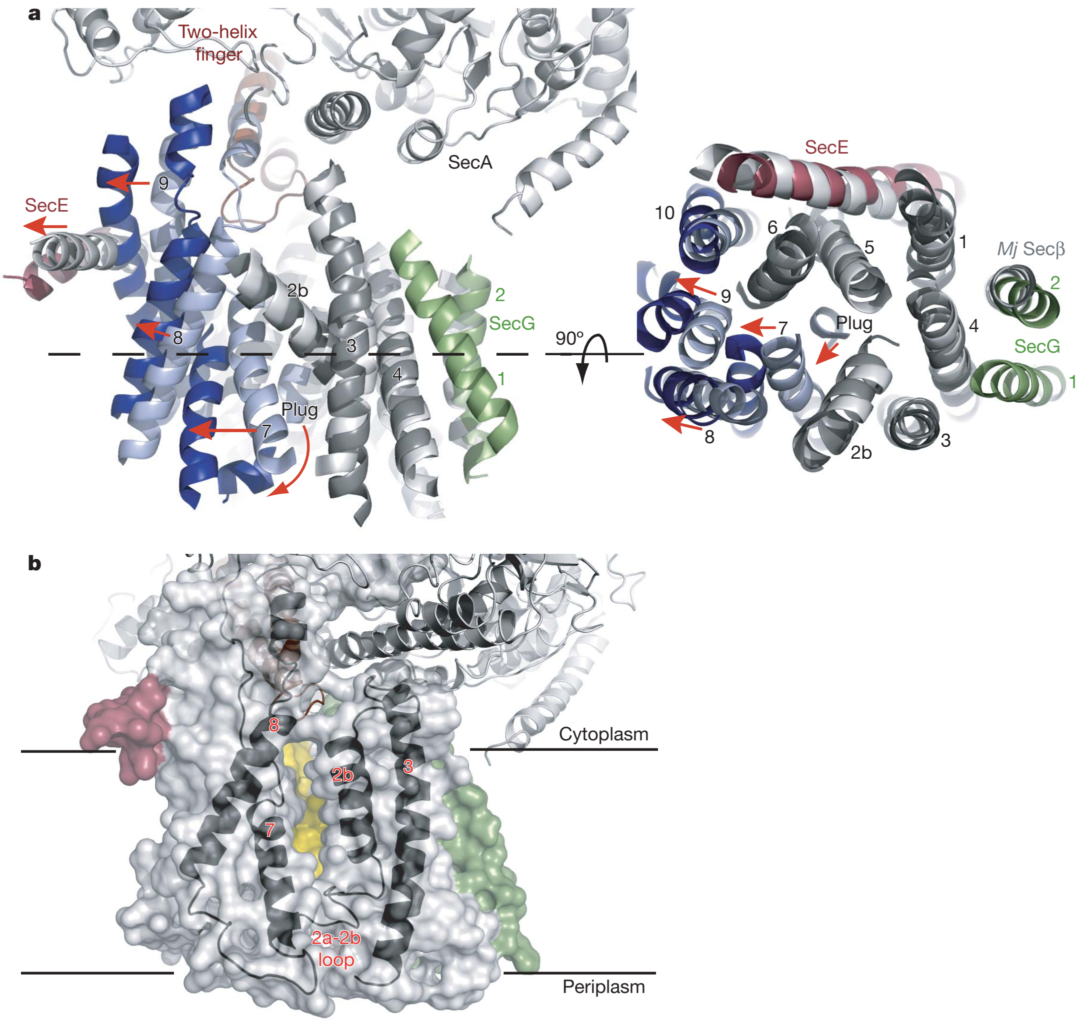


Figure 3 | Conformational changes of SecY. **a**, Comparison of the *T. maritima* SecA–SecYEG complex with the SecY channel from *M. jannaschii* (*Mj*). The alignment is on the basis of TM2b–TM5 of SecY. The view on the right corresponds to a slice through the SecY channel (the plane is indicated by the dashed line on the left). TMs are numbered, and loops are omitted, for clarity. Similarly positioned (root mean squared deviation (r.m.s.d.) ≤1.8 Å) and deviating TMs of *T. maritima* SecY are shown in grey and blue, respectively. Helix movements are indicated by red arrows. **b**, Semi-transparent surface of the SecY channel. The window in the lateral gate of SecY is indicated by a view on the back wall of the channel, formed by TM5 and TM10 (in yellow). TM2b, TM3, TM7 and TM8 are shown as black cartoons.

SecA binds to the SecY complex, leading to a pre-activated state that must prepare the channel for the arrival of the substrate; this state is probably similar to the one represented by our structure. Finally, the polypeptide chain inserts into the SecY channel as a loop, with the signal sequence intercalated into the lateral gate between TM2b and TM7, and the following region in the pore proper[10]. Although our structure lacks a translocating polypeptide, it suggests how each of these different steps could occur.

Capturing a polypeptide by SecA. The marked rotation of the PPXD (Fig. 2a–c) suggests that SecA captures its substrates in a clamp formed by the PPXD, NBD2 and parts of the HSD (Fig. 4a). Spin-label perturbation experiments provide evidence that a polypeptide chain indeed binds to SecA's clamp[34]. The inside of the clamp is lined with conserved residues and contains a hydrophobic patch (Fig. 4a, b). The membrane-proximal part of the clamp is formed by the two-helix finger, and its back is formed by the strictly conserved β-strands connecting NBD1 with the PPXD (shown in orange in Fig. 2a–c). The rotation of the PPXD may allow SecA to bind to a substrate in the cytosol before the complex associates with SecY[34]. Once bound, the clamp would position the polypeptide chain right above the SecY pore (Figs 1b, c and 4c). In our structure, the closed clamp barely leaves enough space to accommodate an extended polypeptide chain, but the loops of the PPXD could move in the presence of a translocation substrate.

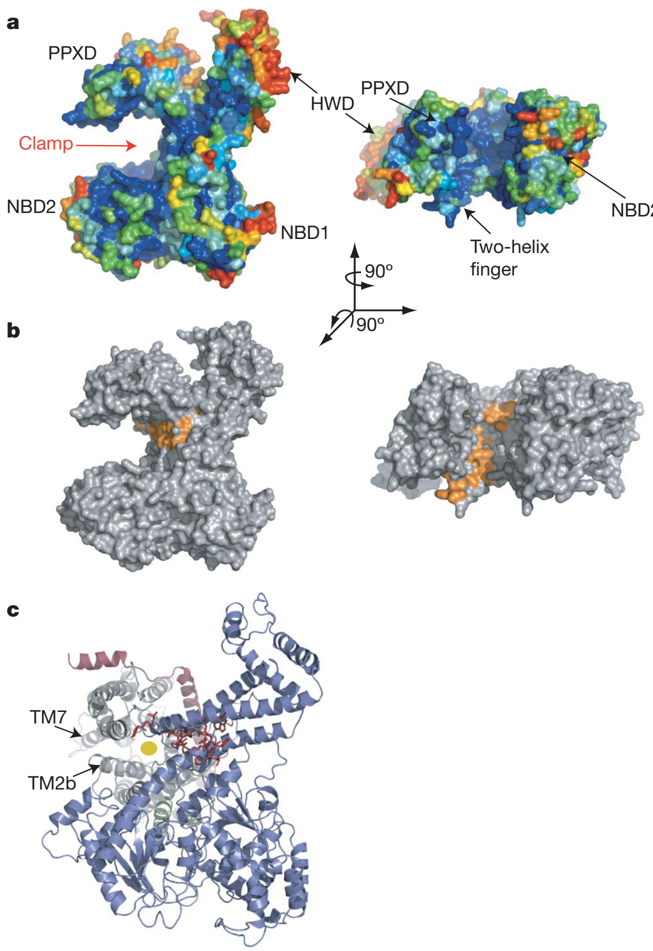


Figure 4 | The polypeptide clamp of SecA. **a**, Surface representation of *B. subtilis* SecA (PDB accession 1TF2) with conserved amino acid residues colour coded (blue denotes the highest, and red denotes the lowest conservation). **b**, As in **a**, except that the conserved hydrophobic residues along the proposed polypeptide clamp of SecA are shown in orange. **c**, The *T. maritima* SecA–SecYEG complex with SecA shown in blue, and SecY, SecE and SecG shown in grey, red and green, respectively. The PPXD was removed for clarity. The residues that were labelled orange in **b** are shown as red sticks. The yellow circle indicates the translocation pore in SecY.

A recent NMR study reported that a synthetic signal peptide binds to the outside of the clamp (Supplementary Fig. 10)[36], but the hydrophobic nature of the interior of the clamp suggests an additional or alternative binding site for signal sequences. A shared binding site inside the clamp for both signal sequences and downstream sequences would explain why the transfer of the signal sequence from SecA into the SecY channel requires a full cycle of ATP hydrolysis as for any other polypeptide segment[6], and why SecA can translocate polypeptides lacking a signal sequence when associating with a signal-sequence suppressor mutant of SecY (PrlA)[37].

The cytosolic chaperone SecB interacts with the C-terminal domain of SecA[38], which is missing in our structures. However, the C terminus of SecA is located on the membrane-distal side, indicating that SecB and SecY bind to opposite sides of the flat SecA molecule. This arrangement would allow the stepwise transfer of a polypeptide from SecB to SecA to SecY[35].

Signal-sequence binding by SecY. Once released from SecA, the signal sequence of a translocation substrate intercalates into the lateral gate of SecY, with the hydrophobic region forming a helix of approximately two turns (Fig. 5a)[10]. The SecA-induced window in the lateral gate of SecY could accommodate this helix after only a slight further expansion. The signal sequence would insert as a loop from the inside of the cytoplasmic funnel of SecY. The positively charged N terminus might be retained on the cytosolic side of the window by an interaction with negatively charged phospholipid head groups. Intercalation of the signal sequence would initiate opening of the channel by releasing the plug from the position observed in our structure towards the periplasmic space. Insertion of the polypeptide segment after the signal sequence into the pore proper would then fix the open state.

Assuming that a window in the lateral gate is maintained even when the plug is fully displaced, a polypeptide segment inside the pore could contact the hydrocarbon chains of the surrounding lipids. If a segment is sufficiently hydrophobic, it would partition into the lipid phase, becoming a TM segment of a membrane protein[39]. Seams of the lateral gate on either side of the membrane would prevent phospholipid head groups from moving into the pore, explaining how the channel can be stable despite opening towards the centre of the lipid bilayer.

Movement of the polypeptide by SecA. We propose that the tip of the two-helix finger of SecA moves up and down inside the cytoplasmic funnel of SecY during the ATP hydrolysis cycle, thereby pushing the polypeptide chain into the pore (Fig. 5b). The fingertip indeed contacts a translocating polypeptide[40] and its mutation or deletion abolishes translocation[40–42]. The finger would move the polypeptide when SecA binds to ATP, and it would release it and reset on ATP hydrolysis. This mechanism is analogous to that proposed for hexameric RecA-like ATPases, such as ClpX, ClpA, HslU and p97 (refs 43–47). In these cases, each monomer has a loop with an essential Tyr (or Trp in the case of p97) at its tip, which moves inside the central pore and contacts the polypeptide. Most SecAs also have a Tyr in the loop at the tip of the two-helix finger, which is essential for translocation[40]. The movement of the two-helix finger could also cause the cyclic widening and constriction of the SecY pore. Our structure makes it unlikely that any part of SecA crosses the channel during translocation[13].

Coordinated with movements of the two-helix finger, the clamp of SecA would hold the translocating polypeptide chain in the ADP state and release it in the ATP state (Fig. 5b). The clamp may just tighten and widen during the ATP hydrolysis cycle, without a large-scale rotation of the PPXD. A translocating polypeptide inside the clamp might transiently form a short β-sheet with the two β-strands connecting the NBD1 with the PPXD. This sequence-independent mode of interaction has been observed in some SecA crystal structures, in which either the weakly structured C-terminal tail of SecA or a segment of a neighbouring SecA molecule forms the extra β-strand[14,22].

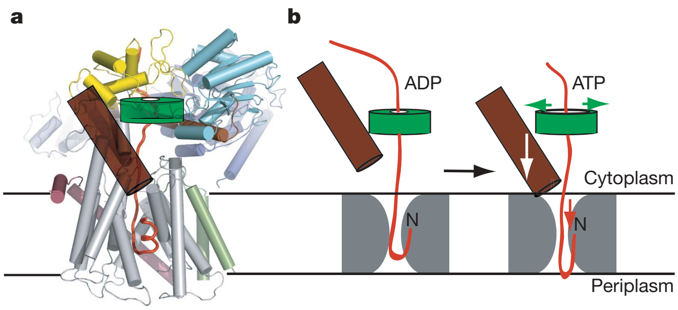


Figure 5 | Model for SecA-mediated protein translocation. **a**, Proposed polypeptide translocation pathway through the SecA–SecYEG complex. The signal sequence of a hypothetical polypeptide (red trace) is intercalated into the lateral gate of SecY. The clamp of SecA is highlighted as a semi-transparent green torus, and the two-helix finger is shown as a brown cylinder. **b**, Upon ATP binding, the clamp of SecA would widen, allowing the two-helix finger to bind to the polypeptide and move it into the channel. After ATP hydrolysis, the clamp would tighten and the finger would reset. SecA probably adopts more than two conformational states during ATP hydrolysis. The structure might represent a situation in which the two-helix finger is in its 'down-state' and the clamp is already closed, although the conformation of the clamp may be different in the presence of a translocating polypeptide.

The formation of a short β-sheet with the translocating polypeptide chain could restrict movements of the PPXD, thus stabilizing its interaction with the loops of SecY. At the end of translocation, when the polypeptide chain has moved past the β-strands, the PPXD would be free to rotate away from the NBD2, resulting in the release of the polypeptide from the clamp and the disengagement of SecA from the SecY channel. The perfect alignment of the mouth of the clamp with the lateral gate of the SecY channel (Supplementary Fig. 9) suggests that a polypeptide chain could laterally exit both SecA and SecY at the same time, an event that may be required when SecA is involved in the integration of membrane proteins[2–4].

The postulated movements of the PPXD and the two-helix finger must originate from the relative movement of the NBDs during the nucleotide hydrolysis cycle. Although we can only speculate about the exact mechanism by which the conformational changes are propagated (see Supplementary Discussion), it is probable that they involve both the loop of the PPXD that interacts with the NBDs and the long helix of the HSD[48] (Figs 1c and 2c). Our structure now provides the basis for determining the mechanistic details of this process.

METHODS SUMMARY

All SecA ATPases and SecY channels were individually expressed in *E. coli* and purified by metal affinity chromatography, gel filtration and ion exchange chromatography. The detergents dodecyl-β-D-maltopyranoside and Cymal-6 were used for the purification of the SecY channels. SecA–SecY complexes were formed by dialysis against a buffer containing ADP–BeF_x and were crystallized by the vapour diffusion technique. After freezing in liquid nitrogen, diffraction data were collected at synchrotron beamlines ID-19 and 24ID at the Argonne National Laboratory, and beamline X29 at the Brookhaven National Laboratory. Molecular replacement was used to locate *B. subtilis* SecA in its complex with *T. maritima* SecYE. A preliminary model of this complex was then used for molecular replacement with a complex of *T. maritima* SecA–SecYEG. The phases were improved by non-crystallographic symmetry (NCS) and cross-crystal averaging. In addition, experimental phases for the *T. maritima* complex were obtained from Se-Met derivatized SecY channel. The final model was refined to a R_{work} and R_{free} factor of 27.9% and 30.3%, respectively, and was used to solve the structure of the *B. subtilis*–*A. aeolicus* complex by molecular replacement. A detailed outline of the experimental procedure is given in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 June; accepted 27 July 2008.

1. Rapoport, T. A. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* **450**, 663–669 (2007).
2. Neumann-Haefelin, C., Schafer, U., Muller, M. & Koch, H. G. SRP-dependent co-translational targeting and SecA-dependent translocation analyzed as individual steps in the export of a bacterial protein. *EMBO J.* **19**, 6419–6426 (2000).
3. Qi, H. Y. & Bernstein, H. D. SecA is required for the insertion of inner membrane proteins targeted by the *Escherichia coli* signal recognition particle. *J. Biol. Chem.* **274**, 8993–8997 (1999).
4. Duong, F. & Wickner, W. Sec-dependent membrane protein biogenesis: SecYEG, preprotein hydrophobicity and translocation kinetics control the stop-transfer function. *EMBO J.* **17**, 696–705 (1998).
5. van den Berg, B. *et al.* X-ray structure of a protein-conducting channel. *Nature* **427**, 36–44 (2004).
6. Osborne, A. R. & Rapoport, T. A. Protein translocation is mediated by oligomers of the SecY complex with one SecY copy forming the channel. *Cell* **129**, 97–110 (2007).
7. Duong, F. Binding, activation and dissociation of the dimeric SecA ATPase at the dimeric SecYEG translocase. *EMBO J.* **22**, 4375–4384 (2003).
8. Harris, C. R. & Silhavy, T. J. Mapping an interface of SecY (PrlA) and SecE (PrlG) by using synthetic phenotypes and *in vivo* cross-linking. *J. Bacteriol.* **181**, 3438–3444 (1999).
9. Tam, P. C., Maillard, A. P., Chan, K. K. & Duong, F. Investigating the SecY plug movement at the SecYEG translocation channel. *EMBO J.* **24**, 3380–3388 (2005).
10. Plath, K., Mothes, W., Wilkinson, B. M., Stirling, C. J. & Rapoport, T. A. Signal sequence recognition in posttranslational protein transport across the yeast ER membrane. *Cell* **94**, 795–807 (1998).
11. Brundage, L., Hendrick, J. P., Schiebel, E., Driessen, A. J. M. & Wickner, W. The purified *E. coli* integral membrane protein SecY/E is sufficient for reconstitution of SecA-dependent precursor proteintranslocation. *Cell* **62**, 649–657 (1990).
12. Akimaru, J., Matsuyama, S. I., Tokuda, H. & Mizushima, S. Reconstitution of a protein translocation system containing purified SecY, SecE, and SecA from *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **88**, 6545–6549 (1991).
13. Economou, A. & Wickner, W. SecA promotes preprotein translocation by undergoing ATP-driven cycles of membrane insertion and deinsertion. *Cell* **78**, 835–843 (1994).
14. Hunt, J. F. *et al.* Nucleotide control of interdomain interactions in the conformational reaction cycle of SecA. *Science* **297**, 2018–2026 (2002).
15. Or, E., Navon, A. & Rapoport, T. Dissociation of the dimeric SecA ATPase during protein translocation across the bacterial membrane. *EMBO J.* **21**, 4470–4479 (2002).
16. Or, E., Boyd, D., Gon, S., Beckwith, J. & Rapoport, T. The bacterial ATPase SecA functions as a monomer in protein translocation. *J. Biol. Chem.* **280**, 9097–9105 (2005).
17. Alami, M., Dalal, K., Lelj-Garolla, B., Sligar, S. G. & Duong, F. Nanodiscs unravel the interaction between the SecYEG channel and its cytosolic partner SecA. *EMBO J.* **26**, 1995–2004 (2007).
18. Jilaveanu, L. B., Zito, C. R. & Oliver, D. Dimeric SecA is essential for protein translocation. *Proc. Natl Acad. Sci. USA* **102**, 7511–7516 (2005).
19. de Keyzer, J. *et al.* Covalently dimerized SecA is functional in protein translocation. *J. Biol. Chem.* **280**, 35255–35260 (2005).
20. Mitra, K., Frank, J. & Driessen, A. Co- and post-translational translocation through the protein-conducting channel: analogous mechanisms at work? *Nature Struct. Mol. Biol.* **13**, 957–964 (2006).
21. Osborne, A. R., Clemons, W. M. Jr & Rapoport, T. A. A large conformational change of the translocation ATPase SecA. *Proc. Natl Acad. Sci. USA* **101**, 10937–10942 (2004).
22. Zimmer, J., Li, W. & Rapoport, T. A. A novel dimer interface and conformational changes revealed by an X-ray structure of *B. subtilis* SecA. *J. Mol. Biol.* **364**, 259–265 (2006).
23. Shiba, K., Ito, K., Yura, T. & Cerretti, D. P. A defined mutation in the protein export gene within the spc ribosomal protein operon of *Escherichia coli*: isolation and characterization of a new temperature-sensitive secY mutant. *EMBO J.* **3**, 631–635 (1984).
24. Mori, H. & Ito, K. An essential amino acid residue in the protein translocation channel revealed by targeted random mutagenesis of SecY. *Proc. Natl Acad. Sci. USA* **98**, 5128–5133 (2001).
25. Chiba, K., Mori, H. & Ito, K. Roles of the C-terminal end of SecY in protein translocation and viability of *Escherichia coli*. *J. Bacteriol.* **184**, 2243–2250 (2002).
26. Mori, H., Shimizu, Y. & Ito, K. Superactive SecY variants that fulfill the essential translocation function with a reduced cellular quantity. *J. Biol. Chem.* **277**, 48550–48557 (2002).
27. de Vrije, T., de Swart, R., Dowhan, W., Tommassen, J. & de Kruijff, B. Phosphatidylglycerol is involved in protein translocation across *Escherichia coli* inner membranes. *Nature* **334**, 173–175 (1988).
28. Lill, R., Dowhan, W. & Wickner, W. The ATPase activity of SecA is regulated by acidic phospholipids, SecY, and the leader and mature domains of precursor proteins. *Cell* **60**, 271–280 (1990).
29. Velankar, S. S., Soultanas, P., Dillingham, M. S., Subramanya, H. S. & Wigley, D. B. Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* **97**, 75–84 (1999).
30. Nishiyama, K., Suzuki, T. & Tokuda, H. Inversion of the membrane topology of SecG coupled with SecA-dependent preprotein translocation. *Cell* **85**, 71–81 (1996).

31. Satoh, Y., Matsumoto, G., Mori, H. & Ito, K. Nearest neighbor analysis of the SecYEG complex. 1. Identification of a SecY–SecG interface. *Biochemistry* **42,** 7434–7441 (2003).
32. Breyton, C., Haase, W., Rapoport, T. A., Kuhlbrandt, W. & Collinson, I. Three-dimensional structure of the bacterial protein-translocation complex SecYEG. *Nature* **418,** 662–665 (2002).
33. Bostina, M., Mohsin, B., Kuhlbrandt, W. & Collinson, I. Atomic model of the *E. coli* membrane-bound protein translocation complex SecYEG. *J. Mol. Biol.* **352,** 1035–1043 (2005).
34. Cooper, D. B., Smith, V. F., Crane, J. M., Roth, H. C., Lilly, A. A. & Randall, L. L. SecA, the motor of the secretion machine, binds diverse partners on one interactive surface. *J. Mol. Biol.* **382,** 74–87 (2008).
35. Hartl, F. U., Lecker, S., Schiebel, E., Hendrick, J. P. & Wickner, W. The binding cascade of SecB to SecA to SecY/E mediates preprotein targeting to the *E. coli* plasma membrane. *Cell* **63,** 269–279 (1990).
36. Gelis, I. *et al.* Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. *Cell* **131,** 756–769 (2007).
37. Prinz, W. A., Spiess, C., Ehrmann, M., Schierle, C. & Beckwith, J. Targeting of signal sequenceless proteins for export in *Escherichia coli* with altered protein translocase. *EMBO J.* **15,** 5209–5217 (1996).
38. Zhou, J. & Xu, Z. Structural determinants of SecB recognition by SecA in bacterial protein translocation. *Nature Struct. Biol.* **10,** 942–947 (2003).
39. Heinrich, S. U., Mothes, W., Brunner, J. & Rapoport, T. A. The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain. *Cell* **102,** 233–244 (2000).
40. Erlandson, K. J. *et al.* A role for the two-helix finger of the SecA ATPase in protein translocation. *Nature* doi:10.1038/nature07439 (this issue).
41. Jarosik, G. P. & Oliver, D. B. Isolation and analysis of dominant SecA mutations in *Escherichia coli*. *J. Bacteriol.* **173,** 860–868 (1991).
42. Karamanou, S. *et al.* A molecular switch in SecA protein couples ATP hydrolysis to protein translocation. *Mol. Microbiol.* **34,** 1133–1145 (1999).
43. Wang, J. *et al.* Crystal structures of the HslVU peptidase–ATPase complex reveal an ATP-dependent proteolysis mechanism. *Structure* **9,** 177–184 (2001).
44. Siddiqui, S. M., Sauer, R. T. & Baker, T. A. Role of the processing pore of the ClpX AAA+ ATPase in the recognition and engagement of specific protein substrates. *Genes Dev.* **18,** 369–374 (2004).
45. Hinnerwisch, J., Fenton, W. A., Furtak, K. J., Farr, G. W. & Horwich, A. L. Loops in the central channel of ClpA chaperone mediate protein binding, unfolding, and translocation. *Cell* **121,** 1029–1041 (2005).
46. DeLaBarre, B., Christianson, J. C., Kopito, R. R. & Brunger, A. T. Central pore residues mediate the p97/VCP activity required for ERAD. *Mol. Cell* **22,** 451–462 (2006).
47. Martin, A., Baker, T. A. & Sauer, R. T. Diverse pore loops of the AAA+ ClpX machine mediate unassisted and adaptor-dependent recognition of ssrA-tagged substrates. *Mol. Cell* **29,** 441–450 (2008).
48. Mori, H. & Ito, K. The long α-helix of SecA is important for the ATPase coupling of translocation. *J. Biol. Chem.* **281,** 36249–36256 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. van den Berg and P. Bendapudi for initial experiments with the *B. subtilis* SecA and *T. maritima* SecY complex, G. Skiniotis for electron microscopy analysis, W. Li for his help with data processing, the staff at Advanced Photon Source beamlines ID-19 and 24ID and at Brookhaven National Laboratory beamline X29, and the SBGrid consortium at Harvard Medical School. We thank A. Brunger and S. Harrison for comments, and A. Brunger, B. van den Berg, W. Li, A. Osborne and S. Schulman for critical reading of the manuscript. The work was supported by a National Institutes of Health grant. T.A.R. is an HHMI investigator. Y.N. is supported by the Damon Runyon Cancer Research Foundation (DRG-#1953-07).

Author Information Coordinates for the *T. maritima* SecA–SecYEG complex have been deposited in the RSCB Protein Data Bank (PDB) under the accession code 3DIN. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to T.A.R. (tom_rapoport@hms.harvard.edu).

METHODS

Protein purification. A fragment of *B. subtilis* SecA comprising residues 2–780 was expressed with a cleavable N-terminal His_6-tag from an isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible promoter, as described[21].

T. maritima SecA (residues 1–816) was expressed in *E. coli* Bl21DE3 as a C-terminal His_6-tagged protein under an IPTG-inducible promoter. After cell lysis in a microfluidizer in HS buffer (20 mM Tris, pH 7.5, 1 M NaCl, 1 mM Tris(2-carboxyethyl) phosphine hydrochloride (TCEP)), the extract was centrifuged for 30 min at 205,000*g* in a Beckman Ti45 rotor at 4 °C. The supernatant was loaded onto a Ni-NTA Sepharose column (Qiagen) and washed with 200 ml HS buffer containing 20 mM imidazole, followed by 200 ml HS buffer containing 20 mM imidazole and 6 M guanidine-HCl. After elution in 50 ml of HS buffer containing 200 mM imidazole and 6 M guanidine-HCl, SecA was refolded by dialysis against 2 l D buffer (20 mM Tris, pH 7.5, 0.5 M NaCl, 5 mM dithiothreitol) overnight at 4 °C. The protein was concentrated and passed over a Superdex S200 gel filtration column (GE Healthcare) in D buffer at room temperature. Fractions lacking a 30 kDa proteolytic fragment of SecA were pooled and concentrated (Centricon) to a final concentration of 100 μM.

The genes coding for *T. maritima* SecY/SecE and SecY/SecE/SecG were cloned into the pBAD22 vector under the arabinose-inducible promoter, as described for the SecYEβ construct from *M. jannaschii*[5], each with a separate translation initiation site. The *T. maritima* SecG sequence is incorrectly annotated in the database; the correct sequence is given in Supplementary Fig. 11. The gene was also cloned into the pACYC vector (Fermentas) under an IPTG-inducible promoter and expressed together with SecYEG in C43(DE3) cells (Lucigen). Approximately 30 g of cells were resuspended in 200 ml R buffer (20 mM Tris, pH 7.5, 200 mM NaCl, 10% (v/v) glycerol) and lysed at 4 °C in a microfluidizer. The membrane fraction was solubilized for 45 min at 4 °C in 150 ml R buffer containing 30 mM dodecyl-β-D-maltopyranoside (Anatrace). The extract was loaded onto a HighPrep-S cation exchange chromatography column (GE Healthcare). The column was washed with R buffer supplemented with 0.6 mM dodecyl-β-D-maltopyranoside and the SecY complex was eluted in R buffer containing 1 M NaCl. The eluate was gel filtered on a Superdex S200 column equilibrated in GF buffer (20 mM Tris, pH 7.5, 100 mM NaCl, 10% (v/v) glycerol, 0.6 mM dodecyl-β-D-maltopyranoside). The peak fractions were loaded onto a Resource S column (GE Healthcare) equilibrated in GF buffer containing 1 mM Cymal-6 (Anatrace). The bound protein was eluted in a linear gradient from 0.1 to 1 M NaCl.

Complexes between the SecA and SecY-channels were formed at a 1:1.2 molar ratio (on the basis of ultraviolet absorbance at 280 nm) and dialysis overnight against a buffer containing 20 mM MES, pH 6.5, 0.1 M NaCl, 10 mM $MgCl_2$, 1 mM ADP, 2 mM $BeCl_2$, 8 mM NaF, 5 mM dithiothreitol, 10% (v/v) glycerol and 1 mM Cymal-6 at 4 °C. The samples were concentrated to 100 μM final concentration and centrifuged for 10 min at 100,000*g* in a Beckman tabletop ultracentrifuge before crystallization.

Se-Met-derivatized *T. maritima* SecYEG was purified as before from cells grown in M9 minimal medium in the presence of 0.2% glucose and 5% (v/v) glycerol. All buffers were supplemented with 5 mM dithiothreitol.

SecYE and C-terminally His-tagged SecG from *Aquifex aeliocus* were expressed in C43(DE3) cells from separate plasmids (pETDuet-1 and pCDFDuet-1, respectively; Novagen) under an IPTG-inducible T7 promoter. Cells from a 12 l culture were treated as described for *T. maritima* SecYEG with the exception that 20 mM decyl-β-D-maltopyranoside (Anatrace) was used for solubilization. The solubilized protein was purified on Ni-NTA agarose (Qiagen) and eluted in 20 mM sodium phosphate, pH 6.0, 150 mM NaCl, 10% glycerol and 1 mM Cymal-6. The eluted protein was further purified by cation exchange chromatography (HiTrapS, GE Healthcare) and gel filtration in R buffer containing 0.5 mM Cymal-6. Complexes between *B. subtilis* SecA and *A. aeolicus* SecYEG were prepared as described above and were purified over a gel filtration column before crystallization.

The biological activity of the purified components was demonstrated by translocation assays with proteoliposomes containing purified *T. maritima* SecYEG, *E. coli* proOmpA, and either *B. subtilis* or *T. maritima* SecA (S. Miller, Y. N. and J. Z., unpublished observations).

Crystallization. Crystallization of the complexes of *T. maritima* SecA–SecYEG and of *B. subtilis* SecA–*T. maritima* SecYE occurred under identical conditions using the sitting drop vapour diffusion technique. Two microlitres of the protein sample was mixed with an equal volume of well solution containing 19–23% (w/v) PEG3350, 50–300 mM $(NH_4)_2SO_4$ and either 0.1 M HEPES, pH 7.5, or 0.1 M Tris, pH 8.5. Crystals grew at 21 °C to approximately 200 × 30 × 30 μm. Crystals of the complexes of *A. aeolicus* SecYEG and *B. subtilis* SecA were obtained using hanging drop vapour diffusion at 21 °C, by mixing equal volumes of protein (10 mg ml^{-1}) and reservoir solution containing 26–28% (w/v) PEG 3350 and 250 mM Li_2SO_4. Crystals grew to a maximum size of about 50 × 50 × 10 μm. All crystals were gradually equilibrated with well solution containing 20% glycerol (v/v) to allow freezing in liquid nitrogen.

Data collection, structure determination and refinement. X-ray diffraction data were collected at 100K at beamlines ID-19 and 24ID of the Argonne National Laboratory, as well as at beamline X29 of the BNL. Single wavelength anomalous diffraction data sets were collected at the Se-absorption edge ($\lambda = 0.9795$) at the BNL beamline X29. The data collection statistics for all crystal forms are listed in Table 1.

Diffraction data were indexed, integrated and scaled using Denzo and Scalepack in HKL2000 (ref. 49). The *B. subtilis* SecA crystal structure (PDB accession 1TF2) lacking the PPXD was used as a search model for molecular replacement of the *B. subtilis*–*T. maritima* SecYE complex data set using Phaser[50,51]. The molecular replacement solution was used to calculate a $2F_o - F_c$ map in CNS 1.2 (ref. 52), allowing to place partial models for TM1, TM3 and TM8–TM10 of SecY, and the SecE subunit and the PPXD of SecA. These TMs, as well as SecA's NBD1 and NBD2, PPXD, HSD and HWD, were modelled as poly-Ala and refined as rigid bodies in CNS 1.2. The model phases were improved by twofold NCS averaging, followed by manual refinement of the placed TMs. The partial model was used as a molecular replacement search model for the *T. maritima* SecA–SecYEG complex data sets. Two copies were placed into the asymmetric units, consistent with self-rotation analysis performed with POLARRFN[53]. The model phases were improved by NCS and cross-crystal averaging in DMMULTI[54,55]using the $P2_1$ and $P2_12_12_1$ data sets of *T. maritima* SecA–SecYEG and the $P6_5$ data set of *B. subtilis* SecA–*T. maritima* SecYE. Furthermore, anomalous difference Fourier analysis allowed the localization of 11 of the 13 Se-Met positions in the derivatized *T. maritima* SecYEG channel (Supplementary Fig. 1). Experimental phases were calculated from the identified selenium positions in a resolution range from 15 to 6.0 Å using SOLVE[56] and were improved by NCS averaging with phase extension to 4.5 Å in the program DM[54]. The experimental electron density map was of sufficient quality (Supplementary Fig. 12) to refine the existing model and to place all remaining TMs of *T. maritima* SecYEG as well as missing regions of SecA. Model building with the program O[57] was guided by the *M. jannaschii* SecYEβ and *B. subtilis* SecA crystal structures (PDB accessions 1RHZ and 1TF2), taking into account sequence homology and interactions between conserved residues. Subsequent electron density maps were calculated after combining model and experimental phases using SIGMAA[58], NCS and cross-crystal averaging, and B-factor sharpening (−20 to −120). The model was refined in CNS 1.2 in the resolution range from 15 to 4.5 Å after applying strict NCS constraints and restraining hydrogen bonds for secondary structure elements[59,60]. Intermediate temperature (1,500 K) simulated annealing and torsion angle refinement were performed, followed by B-factor refinement using group B-factors for backbone and side-chain atoms. The model comprised residues 1–816 for SecA, 8–426 for SecY, 10–65 for SecE and 9–73 for SecG. Residues 42–54 of SecY could not be traced and were omitted from the final model. The presence of an ADP–BeF_x complex was confirmed in a $F_o - F_c$ difference map calculated with model phases after placing an ADP molecule in the nucleotide-binding pocket (Supplementary Fig. 5). The final 4.5 Å *T. maritima* SecA–SecYEG structure was used to improve the model for the complexes of *B. subtilis* SecA with *T. maritima* SecYE or *A. aeolicus* SecYEG. Rigid body refinement of the individual domains of SecA (NBD1 and NBD2, PPXD, HSD and HWD) as well as the N- and C-terminal halves of SecY (TM1–TM5 and TM6–TM10) was performed in CNS 1.2.

Figures shown in the manuscript were generated using Pymol[61]. The colour scheme of conserved residues shown in Fig. 4a was generated with the program CONSURF[62,63].

49. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
50. Storoni, L. C., McCoy, A. J. & Read, R. J. Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D* **60,** 432–438 (2004).
51. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. Likelihood-enhanced fast translation functions. *Acta Crystallogr. D* **61,** 458–464 (2005).
52. Brunger, A. T. Version 1.2 of the Crystallography and NMR system. *Nature Protoc.* **2,** 2728–2733 (2007).
53. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
54. Cowtan, K. D. & Main, P. Phase combination and cross validation in iterated density-modification calculations. *Acta Crystallogr. D* **52,** 43–48 (1996).
55. Cowtan, K. D. & Zhang, K. Y. Density modification for macromolecular phase improvement. *Prog. Biophys. Mol. Biol.* **72,** 245–270 (1999).
56. Terwilliger, T. & Berendzen, J. Automated MAD and MIR structure determination. *Acta Crystallogr. D* **55,** 849–861 (1999).
57. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).

58. Read, R. J. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A* **42**, 140–149 (1986).
59. DeLaBarre, B. & Brunger, A. T. Considerations for the refinement of low-resolution crystal structures. *Acta Crystallogr. D* **62**, 923–932 (2006).
60. Chen, B. *et al.* Determining the structure of an unliganded and fully glycosylated SIV gp120 envelope glycoprotein. *Structure* **13**, 197–211 (2005).
61. DeLano, W. L. The PyMOL Molecular Graphics System. <http://www.pymol.org> (2002).
62. Glaser, F. *et al.* ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* **19**, 163–164 (2003).
63. Landau, M. *et al.* ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33**, W299–302 (2005).

LETTERS

A low-frequency radio halo associated with a cluster of galaxies

G. Brunetti[1], S. Giacintucci[1,2], R. Cassano[1], W. Lane[3], D. Dallacasa[4], T. Venturi[1], N. E. Kassim[3], G. Setti[1,4], W. D. Cotton[5] & M. Markevitch[2]

Clusters of galaxies are the largest gravitationally bound objects in the Universe, containing about 10^{15} solar masses of hot (10^8 K) gas, galaxies and dark matter in a typical volume of 10 Mpc^3. Magnetic fields and relativistic particles are mixed with the gas as revealed by giant 'radio haloes', which arise from diffuse, megaparsec-scale synchrotron radiation at cluster centre[1,2]. Radio haloes require that the emitting electrons are accelerated *in situ* (by turbulence)[3–6], or are injected (as secondary particles) by proton collisions into the intergalactic medium[7–10]. They are found only in a fraction of massive clusters that have complex dynamics[11–14], which suggests a connection between these mechanisms and cluster mergers. Here we report a radio halo at low frequencies associated with the merging cluster Abell 521. This halo has an extremely steep radio spectrum, which implies a high frequency cut-off; this makes the halo difficult to detect with observations at 1.4 GHz (the frequency at which all other known radio haloes have been best studied). The spectrum of the halo is inconsistent with a secondary origin of the relativistic electrons, but instead supports turbulent acceleration, which suggests that many radio haloes in the Universe should emit mainly at low frequencies.

The turbulent re-acceleration model[5,6,15–17] for the origin of giant radio haloes assumes that fossil relativistic particles are re-accelerated by merger-induced turbulence to the energies necessary to produce the observed radio synchrotron emission in relatively weak magnetic fields. The acceleration of fast particles by turbulence, known to be an important process in astrophysics, is due to the resonant scattering of these particles by the turbulent waves (the energy spectrum of the turbulence can be thought of as the superposition of the contribution from turbulent waves with different length scales), leading to a stochastic energization of particles and to the damping of the waves[18]. The connection between mergers and particle acceleration by turbulence in galaxy clusters is complex. It is argued[15,19,20] that it should take place on a timescale of the order of a cluster–subcluster crossing time ($\sim$1 Gyr), during which turbulence is continuously injected on scales of the order of the subcluster size, transported at smaller scales and then dissipated into heating of the intergalactic medium and acceleration of relativistic particles over a fairly large volume. This argument is supported by recent radio observations of a complete sample of X-ray luminous clusters that allow a clear separation between clusters with radio haloes and radio-quiet clusters, suggesting, from the fraction of clusters with radio haloes, that the particle acceleration mechanisms operate sporadically, on timescales $\leq$1 Gyr and in connection with cluster mergers[21].

Abell 521 is an X-ray luminous (8.2×10^{37} W in the 0.1–2.4-keV band) and massive ($\sim 2 \times 10^{15}$ solar masses) galaxy cluster at redshift $z = 0.247$ with ongoing multiple merging episodes[22,23]. Here we report the discovery of a giant radio halo in this cluster by means of deep observations with the Giant Metrewave Radio Telescope (GMRT, India) at 240, 325 and 610 MHz. In Fig. 1a we show the radio image of Abell 521 at 240 MHz, where the radio halo is best imaged. To highlight the diffuse emission, point sources visible in the full-resolution images were subtracted in producing the low-resolution images in Fig. 1. The radio halo is coincident with the cluster X-ray emitting region and correlates with the X-ray emission of the hosting cluster, which is typical of other radio haloes[1,2] (see also Supplementary Information).

Previous higher frequency Very Large Array (VLA) observations at 1,400 MHz did not detect this diffuse emission, instead revealing only the radio relic located on the southeastern boundary of the cluster[23]. The relic coincides with a possible shock front, generated by recent in-fall of a subcluster along the northwest/southeast direction, where relativistic electrons are currently accelerated[24]. Figure 1 clearly shows that the radio halo becomes progressively more dominant over the radio relic at lower frequencies, indicating that its spectrum is much steeper than that of the relic, which has $\alpha \approx 1.5$ (ref. 24; flux is proportional to $\nu^{-\alpha}$, where ν denotes frequency). The patchy structure of the radio halo at 610 MHz (Fig. 1b) indicates the observational difficulty in imaging the emission, the surface brightness of which is already fading at this frequency. The halo disappears between 610 and 1,400 MHz; only an upper limit on the flux of the radio halo at 1,400 MHz can be derived, although faint residual emission in the cluster is still present at this frequency (Fig. 1c and Supplementary Information).

The flux densities of the radio halo at 240, 325 and 610 MHz are plotted in Fig. 2 together with the upper limits at 74 and 1,400 MHz. The upper limits were evaluated by injecting fake radio haloes with different flux densities into the observed data sets, following ref. 21, to estimate the sensitivity of the observations to diffuse emission on the halo length scale. In particular, the upper limit at 74 MHz was derived from VLA Low-frequency Sky Survey data[25], whereas the upper limit at 1,400 MHz was derived from the analysis of the archival VLA data (Supplementary Information). The important result is that the average value of the spectral index, $\alpha \approx 2.1$, is much larger than that of any other known radio halo (typical spectral index is $\alpha \approx 1.2$–1.3 (refs 1, 2)). These extreme spectral properties make Abell 521 a unique system for addressing the origin of the emitting particles in radio haloes.

Such an extremely steep spectrum and the downward spectral curvature (Fig. 2) imply a spectral cut-off at high frequency, which is a well-known signature of turbulent acceleration[4–6,16,17,19]. Synchrotron theory implies a corresponding cut-off in the spectrum of the emitting electrons at $E_e \approx 1.4 B_{nT}^{-1/2} (\nu_c/300)^{1/2}$ GeV, where ν_c is

[1]INAF - Istituto di Radioastronomia, Via P. Gobetti 101, I-40129 Bologna, Italy. [2]Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA. [3]Naval Research Laboratory, Code 7213, Washington DC 20375-5320, USA. [4]Dipartimento di Astronomia, Università di Bologna, Via Ranzani 1, I-40127 Bologna, Italy. [5]National Radio Astronomy Observatory, Charlottesville, Virginia 22903-2475, USA.

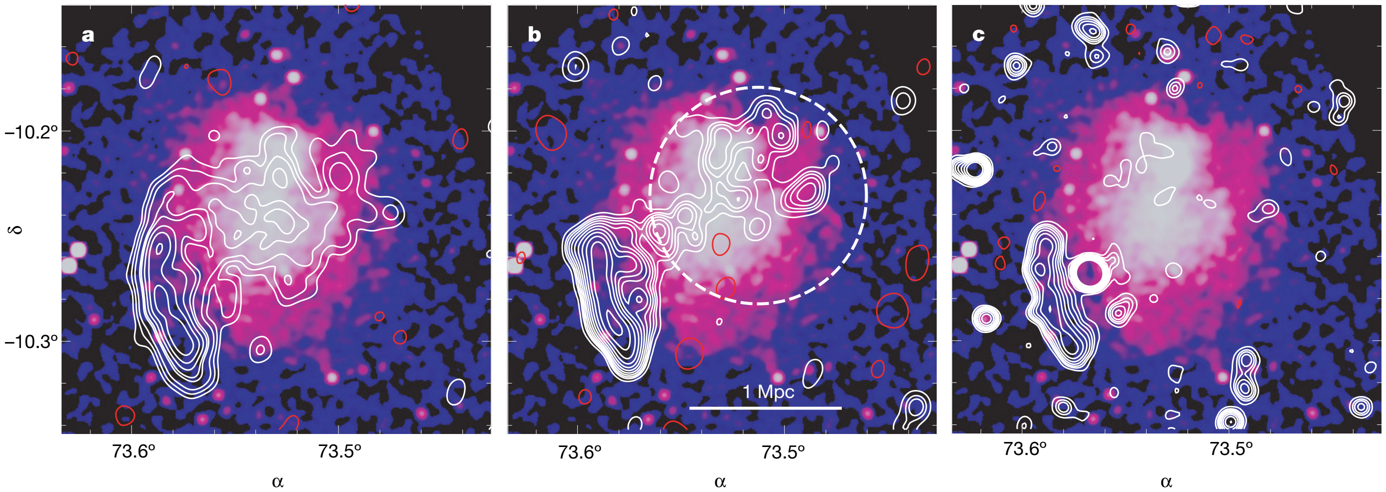


Figure 1 | Radio and X-ray images of Abell 521. Low-resolution radio contours (after subtraction of discrete sources in the field identified at full resolution) overlaid on the Chandra X-ray Observatory 0.5–4-keV X–ray image (corrected for background and exposure and smoothed with a $\sigma = 6$-arcsec Gaussian, as in ref. 24). **a**, GMRT 240-MHz contours at a resolution of 35 arcsec × 35 arcsec. The root-mean-square (r.m.s.) noise level in the cluster region was 220 μJy per beam (full beam resolution used for subtraction of discrete sources was 15.6 arcsec × 12.3 arcsec with an r.m.s. noise level of 190 μJy per beam). Contours are at −0.66 (red), 0.66, 1.3, 2.6, 5.2, 7, 10, 14 and 19 mJy per beam. **b**, GMRT 610-MHz contours at a resolution of 35 arcsec × 35 arcsec. The r.m.s. noise level in the cluster region was 58 μJy per beam (full beam resolution, 9 arcsec × 4 arcsec; r.m.s. noise, 35 μJy per beam). Contours are at −0.17 (red), 0.17, 0.34, 0.48, 0.67, 0.94, 1.3, 1.9, 2.6, 3.7 and 5.2 mJy per beam. The dashed circle indicates the region in which fluxes of the radio halo at the different frequencies (reported in Fig. 2) are measured. **c**, VLA 1.4-GHz contours at a resolution of 25 arcsec × 25 arcsec. Here only the discrete sources within 3 arcmin from the cluster centre were subtracted (other images are presented in Supplementary Information together with a discussion on the subtraction of discrete sources at this frequency). The r.m.s. noise level in the cluster region was 26 μJy per beam (full beam resolution, 12.7 arcsec × 6.9 arcsec; r.m.s. noise, 15 μJy per beam). Contours are at −0.08 (red), 0.08, 0.16, 0.22, 0.32, 0.45, 0.63, 0.89, 1.25, 1.76, 2.48, 3.5 and 4.9 mJy per beam. Coordinate system, J2000.

the cut-off frequency (measured in megahertz) and B_{nT} is the magnetic field (measured in nanotesla). In the framework of the turbulent acceleration scenario, E_{e} pinpoints the energy at which the timescale of

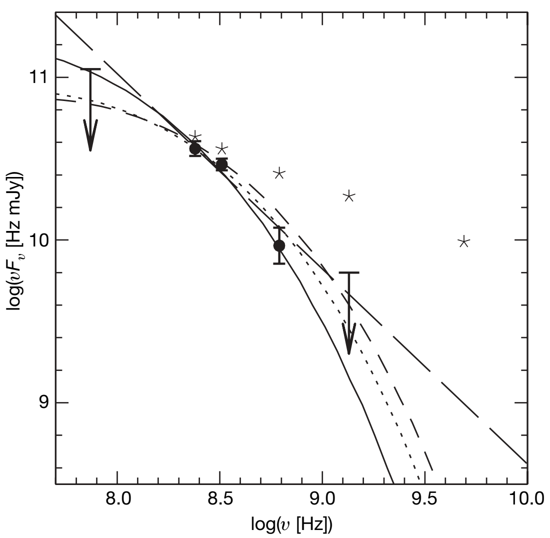


Figure 2 | Spectrum of the radio halo. Data are overlaid on secondary (long-dashed line) and re-acceleration models. Measured fluxes (F_ν) are 152 ± 15 mJy at 240 MHz, 90 ± 7 mJy at 325 MHz, 15 ± 3.5 mJy at 610 MHz, and limits (arrows) are 1.5 Jy at 74 MHz and 5 mJy at 1,400 MHz; uncertainties (and error bars in the figure) are 1 s.d. For comparison, asterisks give the spectral energy distribution of the radio relic (taken from ref. 24). The secondary model assumes that $\delta = 4.2$ and requires that the energy density of relativistic protons is larger than that of the thermal energy (see text). The re-acceleration models assume that 14% of the thermal energy is in magnetosonic waves, with a central value of the magnetic field of $B_{\mathrm{o}} = 0.15$ nT (solid curve); that 14% of the thermal energy is in magnetosonic waves, with $B_{\mathrm{o}} = 0.35$ nT (dotted curve); and that 18% of the thermal energy is in magnetosonic waves, with $B_{\mathrm{o}} = 0.15$ nT (short-dashed curve). All the re-acceleration models adopt a scaling with $B \propto n_{\mathrm{th}}$ (ref. 2 and refs therein).

the electron radiative losses becomes equal to that of the turbulent acceleration. The timescale of electrons emitting at ν_{c} is estimated taking into account the redshift-dependent inverse-Compton losses against the cosmic microwave background and the synchrotron losses[26], as follows, where ν_{c} is again measured in megahertz:

$$\tau \approx 0.95 \frac{B_{\mathrm{nT}}^{1/2}(\nu_{\mathrm{c}}/300)^{-1/2}}{(1+z)^4 + (B_{\mathrm{nT}}/0.32)^2} \text{ Gyr}$$

This means that the electrons responsible for the observed emission should be accelerated on a timescale of ~(1.1–1.4) × 10^8 years, for 0.1–0.5-nT magnetic fields in the radio halo region. Assuming that fast magnetosonic waves are responsible for the acceleration of the emitting particles, following ref. 20 we find that this acceleration efficiency can be achieved under the reasonable assumption that the energy density of these turbulent waves is ~12–18% of the thermal energy (Fig. 2).

Appreciable synchrotron emission can also be produced by secondary electrons injected by collisions between long-lived relativistic protons accumulated in the cluster and the thermal protons in the intergalactic medium, and secondary models have been proposed as alternatives to the re-acceleration model to explain radio haloes[7–10]. The very steep spectral slope of this radio halo rules out secondary models by means of a straightforward energy argument. To explain the spectrum of the radio halo through synchrotron radiation from secondary electrons, the primary protons must have a very steep spectral energy distribution ($N(p) \propto p^{-\delta}$, where $\delta \approx 4.2$ and p denotes the particle momentum). The energy density of relativistic protons, ε_{p}, required to match the observed synchrotron flux through a secondary model can be estimated following the formalism in ref. 27. For an average number density of thermal protons of $n_{\mathrm{th}} \approx 1{,}500\ \mathrm{m}^{-3}$ in the region of the radio halo (consistent with the average thermal density in the same region derived from X-ray observations[22]) and the synchrotron flux measured at 325 MHz, we find that ε_{p} ranges from approximately 3 to 100 times the energy density of the thermal plasma for magnetic field values (averaged in the region of the radio halo) ranging from $B = 0.5$ nT to $B = 0.1$ nT,

respectively. This yields only a lower limit on the energy density of high-energy protons, because for $\delta > 3$ an additional (even dominant) contribution to the energy comes from suprathermal particles with kinetic energies <1 GeV. A secondary origin for the emitting electrons thus implies the unrealistic situation in which clusters are dominated by non-thermal protons. It also violates present upper limits on the energy density of these particles derived from γ-ray observations of several clusters, at the 20% level[28,29].

If the spectrum of these protons has $\delta < 4$, then the synchrotron signal from the secondary electrons produced by proton–proton collisions cannot exceed the 1.4-GHz upper limit in Fig. 2, placing corresponding limits on the energy density of the primary protons. Figure 3 shows upper limits on the energy density of the primary protons in the region of the radio halo as a function of the magnetic field. These limits were obtained following ref. 21 and show that the energy density of relativistic protons in the cluster is less than 1% of the thermal component for $B > 0.2$ nT and $\delta < 2.5$.

A similar conclusion has recently been reached in the analysis of a statistical sample of galaxy clusters without radio haloes[21]. However, in that case a larger energy content of protons was still possible by assuming that clusters without radio haloes have magnetic fields much smaller than those with radio haloes. This alternative (*ad hoc*) possibility can be reasonably ruled out in our case because Abell 521 hosts a radio halo with bolometric radio luminosity comparable to that of classical radio haloes ($\nu P(\nu) \approx 10^{34}$ W). Future observations with the Fermi Gamma-ray Space Telescope (formerly the Gamma-ray Large Area Space Telescope) will reveal galaxy clusters in γ-rays in cases where the energy content of relativistic protons is significantly larger than about 1% of the thermal plasma. The combination of this future data and limits in the radio band (Fig. 3 and ref. 21) may thus provide a powerful tool for constraining the magnetic field strength in galaxy clusters.

As we look at applying the turbulent re-acceleration model to other galaxy clusters, it should be stressed that the maximum energy to which electrons can be re-accelerated and, ultimately, the cut-off frequency in the spectra of radio haloes depend on the level of turbulence and on the properties of the turbulent waves. The spectral cut-off affects our ability to detect radio haloes in the Universe, introducing a strong bias against observing them at frequencies substantially larger than ν_c. Currently known radio haloes are mainly observed at gigahertz frequencies, requiring an efficient turbulent acceleration mechanism. These haloes must result from the rare, most energetic merging events and are therefore hosted only in the most massive, hottest clusters[19,30], in line with observations[11–14]. On the other hand, the majority of radio haloes should form during much more common, but less energetic, merging events, for example between a massive cluster and a substantially smaller subcluster (with mass ratio >5) or between two similar clusters with mass $\leq 10^{15}$ solar masses[19,30]. However, these sources, with a cut-off in the synchrotron spectrum at $\nu_c < 1$ GHz, should be visible only at lower frequencies, because their spectrum should be similar to that of the low-frequency radio halo in Abell 521. Future high-sensitivity radio telescopes operating at low frequencies, such as the Low Frequency Array and the Long Wavelength Array, are expected to discover the majority of these sources and also to test their connection with cluster mergers. At the moment this connection cannot be tested, owing to the lack of observations of samples of galaxy clusters at low radio frequencies.

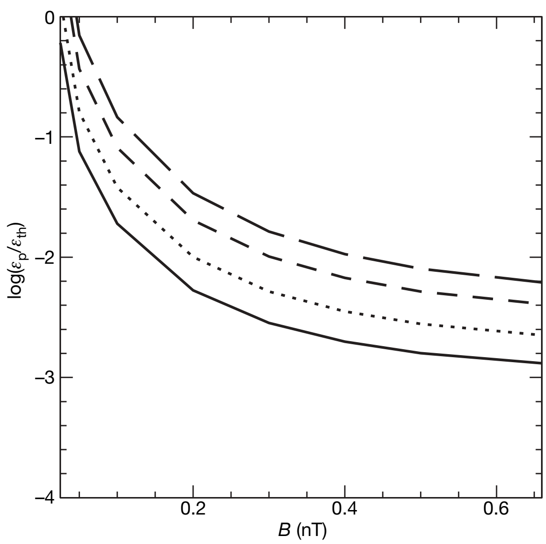


Figure 3 | Limits on the energy density of relativistic protons. Curves give the upper limit on the ratio between the energy density of relativistic protons and that of thermal gas, ε_{th}, as a function of the magnetic field in Abell 521. Calculations are shown assuming that $n_{th} = 1{,}500\ m^{-3}$ and $k_B T = 7$ keV, where T is temperature and k_B is Boltzmann's constant. From bottom to top, the spectral energy distributions of the primary protons have $\delta = 2.3, 2.5, 2.7$ and 2.9 ($N(p) \propto p^{-\delta}$).

Received 29 May; accepted 22 August 2008.

1. Giovannini, G. & Feretti, L. in *Diffuse Radio Sources and Cluster Mergers: Radio Haloes and Relics* (eds Feretti, L., Gioia, I. M. & Giovannini, G.) 197–227 (Astrophys. Space Sci. Library 272, Kluwer, 2002).
2. Ferrari, C., Govoni, F., Schindler, S., Bykov, A. M. & Rephaeli, Y. Observations of extended radio emission in clusters. *Space Sci. Rev.* **134,** 93–118 (2008).
3. Jaffe, W. J. Origin and transport of electrons in the halo radio source in the Coma cluster. *Astrophys. J.* **212,** 1–7 (1977).
4. Schlickeiser, R., Sievers, A. & Thiemann, H. The diffuse radio emission from the Coma cluster. *Astron. Astrophys.* **182,** 21–35 (1987).
5. Brunetti, G., Setti, G., Feretti, L. & Giovannini, G. Particle reacceleration in the Coma cluster: radio properties and hard X-ray emission. *Mon. Not. R. Astron. Soc.* **320,** 365–378 (2001).
6. Petrosian, V. On the nonthermal emission and acceleration of electrons in coma and other clusters of galaxies. *Astrophys. J.* **557,** 560–572 (2001).
7. Dennison, B. Formation of radio halos in clusters of galaxies from cosmic-ray protons. *Astrophys. J.* **239,** L93–L96 (1980).
8. Blasi, P. & Colafrancesco, S. Cosmic rays, radio halos and nonthermal X-ray emission in clusters of galaxies. *Astropart. Phys.* **12,** 169–183 (1999).
9. Dolag, K. & Ensslin, T. A. Radio halos of galaxy clusters from hadronic secondary electron injection in realistic magnetic field configurations. *Astron. Astrophys.* **362,** 151–157 (2000).
10. Pfrommer, C. & Ensslin, T. A. Estimating galaxy cluster magnetic fields by the classical and hadronic minimum energy criterion. *Mon. Not. R. Astron. Soc.* **352,** 76–90 (2004).
11. Giovannini, G., Tordi, M. & Feretti, L. Radio halo and relic candidates from the NRAO VLA Sky Survey. *New Astron.* **4,** 141–155 (1999).
12. Buote, D. A. On the origin of radio halos in galaxy clusters. *Astrophys. J.* **553,** L15–L18 (2001).
13. Cassano, R. *et al.* Revised statistics of radio halos and the reacceleration model. *Astron. Astrophys.* **480,** 687–697 (2008).
14. Venturi, T. *et al.* GMRT Radio Halo Survey in galaxy clusters at $z = 0.2 - 0.4$. II.The eBCS clusters and analysis of the complete sample. *Astron. Astrophys.* **484,** 327–340 (2008).
15. Fujita, Y., Takizawa, M. & Sarazin, C. L. Nonthermal emissions from particles accelerated by turbulence in clusters of galaxies. *Astrophys. J.* **584,** 190–202 (2003).
16. Kuo, P.-H., Hwang, C.-Y. & Ip, W.-H. Diagnostic signatures of radio and hard X-ray emission on particle acceleration processes in the Coma cluster. *Astrophys. J.* **594,** 732–740 (2003).
17. Brunetti, G., Blasi, P., Cassano, R. & Gabici, S. Alfvenic reacceleration of relativistic particles in galaxy clusters: MHD waves, leptons and hadrons. *Mon. Not. R. Astron. Soc.* **350,** 1174–1194 (2004).
18. Melrose, D. B. *Plasma Astrophysics* (Gordon & Breach, 1980).
19. Cassano, R. & Brunetti, G. Cluster mergers and non-thermal phenomena: a statistical magnetoturbulent model. *Mon. Not. R. Astron. Soc.* **357,** 1313–1329 (2005).
20. Brunetti, G. & Lazarian, A. Compressible turbulence in galaxy clusters: physics and stochastic particle re-acceleration. *Mon. Not. R. Astron. Soc.* **378,** 245–275 (2007).
21. Brunetti, G. *et al.* Cosmic rays and radio halos in galaxy clusters: New constraints from radio observations. *Astrophys. J.* **670,** L5–L8 (2007).
22. Arnaud, M., Maurogordato, S., Slezak, E. & Rho, J. A521: A cluster forming at the crossing of two filaments? *Astron. Astrophys.* **355,** 461–478 (2000).
23. Ferrari, C., Arnaud, M., Ettori, S., Maurogordato, S. & Rho, J. Chandra observation of the multiple merger cluster Abell 521. *Astron. Astrophys.* **446,** 417–428 (2006).

24. Giacintucci, S. *et al.* Shock acceleration as origin of the radio relic in A521? *Astron. Astrophys.* **486,** 347–358 (2008).
25. Cohen, A. S. *et al.* The VLA Low-Frequency Sky Survey. *Astron. J.* **134,** 1245–1262 (2007).
26. Sarazin, C. L. The energy spectrum of primary cosmic-ray electrons in clusters of galaxies and inverse Compton emission. *Astrophys. J.* **520,** 529–547 (1999).
27. Brunetti, G. & Blasi, P. Alfvenic reacceleration of relativistic particles in galaxy clusters in the presence of secondary electrons and positrons. *Mon. Not. R. Astron. Soc.* **363,** 1173–1187 (2005).
28. Reimer, O., Pohl, M., Sreekumar, P. & Mattox, J. R. EGRET upper limits on the high-energy gamma-ray emission of galaxy clusters. *Astrophys. J.* **588,** 155–164 (2003).
29. Pfrommer, C. & Ensslin, T. A. Constraining the population of cosmic ray protons in cooling flow clusters with γ-ray and radio observations: Are radio mini-halos of hadronic origin? *Astron. Astrophys.* **413,** 17–36 (2004).
30. Cassano, R., Brunetti, G. & Setti, G. Statistics of giant radio haloes from electron reacceleration models. *Mon. Not. R. Astron. Soc.* **369,** 1577–1595 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge partial support from ASI-INAF I/088/06/0 and PRIN-INAF 2007. G.B. and R.C. acknowledge the Harvard-Smithsonian Center for Astrophysics, the US Naval Research Laboratory and the National Radio Astronomy Observatory for hospitality and partial support during the preparation of the manuscript. Basic research in radio astronomy at the Naval Research Laboratory is supported by 6.1 base funding. The National Radio Astronomy Observatory is operated by Associated Universities, Inc., under cooperative agreement with the National Science Foundation. We thank R. Athreya and the staff of the GMRT for their helping during the observations. The GMRT is run by the National Centre for Radio Astrophysics, Tata Institute of Fundamental Research.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to G.B. (brunetti@ira.inaf.it).

LETTERS

Spontaneous vortices in the formation of Bose–Einstein condensates

Chad N. Weiler[1], Tyler W. Neely[1], David R. Scherer[1], Ashton S. Bradley[2]†, Matthew J. Davis[2] & Brian P. Anderson[1]

Phase transitions are ubiquitous in nature, and can be arranged into universality classes such that systems having unrelated microscopic physics show identical scaling behaviour near the critical point. One prominent universal element of many continuous phase transitions is the spontaneous formation of topological defects during a quench through the critical point[1–3]. The microscopic dynamics of defect formation in such transitions are generally difficult to investigate, particularly for superfluids[4–7]. However, Bose–Einstein condensates (BECs) offer unique experimental and theoretical opportunities for probing these details. Here we present an experimental and theoretical study of the BEC phase transition of a trapped atomic gas, in which we observe and statistically characterize the spontaneous formation of vortices during condensation[8,9]. Using microscopic theories[10–17] that incorporate atomic interactions and quantum and thermal fluctuations of a finite-temperature Bose gas, we simulate condensation and observe vortex formation in close quantitative agreement with our experimental results. Our studies provide further understanding of the development of coherence in superfluids, and may allow for direct investigation of universal phase transition dynamics.

Spontaneous vortex formation in superfluids is intimately connected to superfluid growth. In one model, illustrated in Fig. 1, isolated superfluid regions of characteristic size ξ independently form as the system nears the critical point of the phase transition. These regions with random relative phases merge together during the transition, leading to a continuous phase gradient in the merged fluid. Owing to continuity requirements on the wavefunction, the merging process may trap phase loops of 2π if the merging regions have suitable relative phases, as illustrated in Fig. 1. The superfluid density at the centre of these 2π phase loops is topologically constrained to be zero, resulting in the formation of a quantized vortex; the absence of superfluid at the vortex core may be viewed as arising from destructive interference between merging regions. Although cast here in the context of superfluid growth, spontaneous topological defect formation is a fundamental component of the Kibble–Zurek mechanism[1–3]. Based on universality classes for second-order phase transitions, this mechanism provides a prescription for estimating a correlation length ξ and hence the density of defects, proportional to $1/\xi^2$, that may form. For a continuous phase transition that proceeds quasistatically, ξ diverges at the critical point and therefore no defects are expected. However, in the Kibble–Zurek mechanism the phase transition occurs over a finite time, and the system falls out of equilibrium when the thermalization (or relaxation) rate drops below a quench rate $1/\tau_Q$. At this point ξ is frozen in and essentially remains constant through the critical point. A principal result is that faster quenches lead to an earlier freeze-in time, and hence smaller values of ξ and higher defect densities.

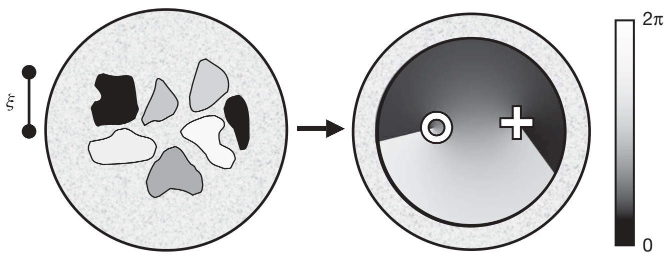


Figure 1 | Schematic of spontaneous vortex formation. Left, as a thermal gas (mottled grey) is cooled through the BEC transition, isolated coherent regions of approximate size ξ and unpredictable phase may form[8,9]. Quantum phase ranges from 0 to 2π, represented here by shades of grey as indicated by the gradient bar at the right. Right, initial coherent regions eventually merge to form a single BEC (continuous greyscale region), potentially forming quantized vortices. Here, a positive (negative) vortex is labelled with a cross (circle), with the phase winding direction corresponding to the direction of superfluid flow and phase gradient about the vortex core.

The Kibble–Zurek mechanism is appealing because of its potential for characterizing a wide variety of phase transitions, irrespective of the microscopic processes involved. A model of condensation in a homogeneous Bose gas describing the transition from a weak-turbulent (kinetic) stage to strong-turbulent (coherent) state has been proposed by Svistunov and co-workers[9,18–21]. In this scheme, as energy is removed from the system the low-energy atomic field modes become macroscopically occupied. Destructive interference between these essentially classical modes leads to nodes in the field, which appear as lines of zero atomic density. Subsequently, a quasicondensate having local coherence but no long-range coherence grows around the lines of zero density, which simultaneously evolve into well-structured vortex cores. Eventually the superfluid relaxes into equilibrium and a true condensate with global phase coherence is achieved. Berloff and Svistunov numerically demonstrated the validity of this picture for the homogeneous Bose gas in simulations of the Gross–Pitaevskii equation[22] (see Supplementary Information). Our work involves an experimental and theoretical exploration of similar phenomena in the condensation of trapped gases. Our approach also has the potential to investigate the relationship of the Kibble–Zurek mechanism to phase transition dynamics of BECs.

In previous work we demonstrated that vortices can form during the controlled merging of three independent BECs with uncorrelated phases[23], an analogue of the Kibble–Zurek mechanism. Here we study vortex formation by evaporatively cooling an atomic gas through the BEC phase transition in a single axially symmetric oblate harmonic trap (see Methods). To probe condensate growth dynamics under varying cooling conditions, we use two temperature quenches: quench A uses a 6-s radiofrequency evaporative cooling

[1]College of Optical Sciences, University of Arizona, Tucson, Arizona 85721, USA. [2]ARC Centre of Excellence for Quantum-Atom Optics, School of Physical Sciences, University of Queensland, Brisbane, Queensland 4072, Australia. †Present address: Jack Dodd Centre for Quantum Technology, Department of Physics, University of Otago, PO Box 56, Dunedin, New Zealand.

ramp, and quench B uses a sudden jump to a final radiofrequency value (see Methods). Plots of temperature and condensate number versus time for both quenches are shown in Fig. 2a. Following Anglin and Zurek[8] (see Supplementary Information), we estimate a correlation length of $\xi \approx 0.6\,\mu$m near the critical point for both quenches. Because ξ is about a factor of 6 smaller than our radial harmonic oscillator length $a_r \approx 3.8\,\mu$m characterizing a condensate radius for small atom numbers, we would not expect global phase coherence at the critical temperature, suggesting that spontaneous vortex formation could occur in our experiments.

To look for vortices, we suddenly remove the trapping potential after the 6-s evaporative cooling ramp of quench A or 1.5 s after the radiofrequency jump for quench B. Each BEC ballistically expands and is then imaged along the vertical direction (the z axis), which coincides with the symmetry axis of the trap. Vortex cores aligned with the z axis appear as holes in the column-density distribution, as shown in Fig. 3a. We emphasize that our procedure does not impart net angular momentum to the atomic cloud, such as through phase engineering[24] or stirring[25]; our observations thus represent a new regime for the study of quantized vortex nucleation in BECs (see Supplementary Information for further discussion).

We simulate condensate formation using the stochastic Gross–Pitaevskii equation (SGPE) formalism[13,14] that describes the highly occupied, low-energy modes of a Bose gas with a classical field. The field evolves according to a generalized Gross–Pitaevskii equation that includes dissipation and thermal noise describing collisions between the partially condensed matter waves and the high-energy atoms in the thermal cloud. Because evaporative cooling is difficult to simulate realistically[26], and the details are often qualitatively

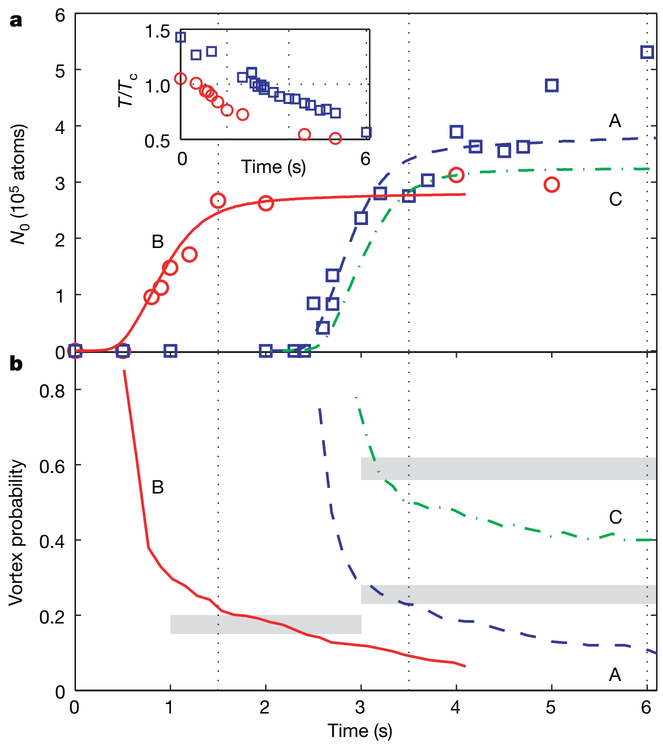


Figure 2 | Condensate formation and vorticity. **a**, Condensate number N_0 versus time. Blue squares (red circles) indicate experimental data for quench A (B), and lines indicate corresponding numerical simulations. The green dot-dashed line is the numerical result for the toroidal trap (quench C). Vertical dotted lines indicate the observation times for which experimental statistics are generated. Inset, experimentally measured temperatures for quenches A and B ($\tau_Q \approx 7$ s and 5 s, respectively). **b**, The probability of finding at least one vortex passing through the $z = 0$ plane plotted for all three simulated quenches. Grey regions indicate the experimental measurement range for each data set.

unimportant, we use an idealized cooling model with a sudden jump in chemical potential and temperature of the thermal cloud through the condensation critical point. This leaves the SGPE classical field out of equilibrium with the thermal cloud; the subsequent return to equilibrium results in condensate formation. Figure 2a shows the growth in condensate number for the simulations of both quenches. In our simulations, the initial and final thermal cloud parameters have been chosen to match the experimental results, and the coupling between the thermal cloud and the classical field is then adjusted to give good agreement with the experimentally observed BEC growth curves. This approach allows a meaningful comparison of other observables such as vortex statistics with the experimental data. Further discussion can be found in the Methods and Supplementary Information.

As shown in Fig. 3b, c, vortices spontaneously form in our simulations, where each realization can be interpreted as the numerical analogue of a single experimental run. We therefore study vortex dynamics in each growing condensate to compare vortex formation statistics with our experimental results. In both our laboratory and numerical procedures, for each quench we repeat the BEC creation procedure and analyse statistics of vortex observations. For each data set described below, we extract the fraction of images showing at least one vortex core within a displacement of $0.8R_{TF}$ from the BEC centre, where R_{TF} is the BEC Thomas–Fermi radius in the $z = 0$ plane[27]. This fraction serves as our estimate of the probability of observing spontaneously formed vortices in a single run.

Because localized decreases in the density profile of an experimentally obtained image may not always clearly indicate the presence of a core (for example, owing to tilting or bending with respect to the z axis) our experimental uncertainty ranges are defined by our ability to determine visually whether an image shows a vortex. For quench A, 23–28% of 90 total images contain at least one visible vortex core. For quench B, 15–20% of 98 total images show at least one core. Although the two quenches use quite different radiofrequency evaporation trajectories, they show similar cooling and BEC growth rates. We can thus expect statistical similarities between the two data sets. Further statistical details, including results of observing multiple cores per image, are given in the Supplementary Information.

From our simulations, we can analyse vortex observation probabilities as a function of time for each quench. To determine the presence of a vortex we consider an instantaneous slice of the classical field in the $z = 0$ plane of the trap, and detect all phase loops of $\pm 2\pi$ within a displacement of $0.8R_{TF}$ from the BEC centre (here, R_{TF} is based on the time-dependent condensate number). We find that the

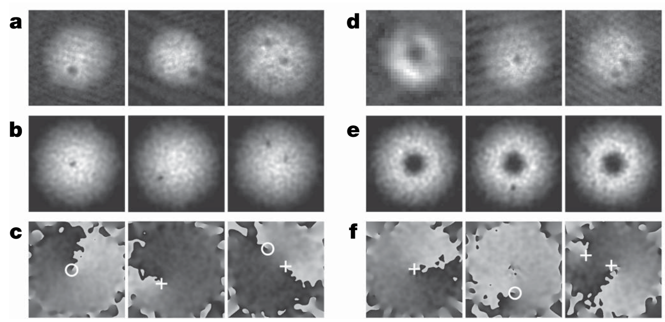


Figure 3 | Vortices in the harmonic and toroidal traps. **a**, Images of BECs created in a harmonic trap, showing single vortices (left, centre) and two vortices (right). Each image is 200 μm square. **b**, **c**, Sample simulation results from quench B, showing in-trap integrated column densities along z (in **b**) and associated phase profiles in the $z = 0$ plane (in **c**), with vortices indicated by crosses and circles at $\pm 2\pi$ phase windings. **d**, Left image, phase-contrast experimental image of a BEC in the toroidal trap. Image is 70 μm square. Remaining images, vortices in 200-μm-square expansion images of BECs created in the toroidal trap. **e**, **f**, Simulations of BEC growth in the toroidal trap show vortices (as in **b**, **c**) and persistent currents.

majority of vortices are aligned with the z axis of the trap. The vortex observation probabilities obtained from ~300 simulation runs for each quench are plotted against time in Fig. 2b, with comparison to the experimental statistics.

According to our simulations, the number of vortices decreases with time, consistent with our model of a thermal bath that is independent of time and has no angular momentum; the thermodynamic final state should therefore be a condensate without vortices. In this respect the simulations diverge from our experimental observations, where there is no significant variation of vortex observation probabilities over a timescale of a few seconds. This low damping rate is consistent with the comparatively small thermal component observed, indicating that a kinetic theory of thermal cloud dynamics may be needed to account fully for the long-time behaviour of the experiment. We therefore compare our simulation results at $t = 3.5$ s for quench A, and $t = 1.5$ s for quench B, based on experimental observations of negligible vortex damping once the BEC is nearly fully formed.

By focusing a blue-detuned laser beam propagating along the z axis into the centre of the trap (see Methods), we experimentally studied BEC growth in a toroidal potential in which a BEC may display both persistent superfluid current[28] about the central barrier as well as free vortices circulating around the barrier. The pinning of superfluid flow may influence both vortex dynamics during BEC growth and observations of vortices after the BEC is formed: a vortex pinned to the barrier reduces the likelihood of complete self-annihilation between pairs of spontaneously formed vortices of opposite charge, thereby increasing the probability of finding a vortex in a BEC. We apply a 6-s final evaporative cooling ramp identical to quench A, and identify this data set as quench C. An *in situ* image of a BEC in the toroidal trap is given in the leftmost image of Fig. 3d. Note that the dark region in the BEC centre indicates atoms displaced by the laser beam; vortices are not visible in this image. After creating each BEC, we ramp down the laser power over 100 ms and immediately thereafter allow the BEC to expand from the trapping potential. For these conditions, we find that 56–62% of 52 images contained at least one visible core; examples are shown in Fig. 3d.

Condensate formation rates were not experimentally measured for quench C; for the simulations we use the parameters of quench A but with an additional repulsive Gaussian barrier. Simulated condensate growth versus time resulted in smaller condensates than for quench A (also observed experimentally) as shown by the green dot-dashed curve in Fig. 2a. Examples of numerically obtained column density and phase are shown in Fig. 3e, f. Vortex observation statistics for 300 runs are plotted as a green dot-dashed line in Fig. 2b; we find that the vortex observation probability is about twice that of the harmonic trap of quench A, as is also the case with the experimental data, but is somewhat lower in overall magnitude than the experimental observations. In contrast to the harmonic case, the curve does not show decay below 40%, which corresponds to the fraction of runs with vortices pinned by the Gaussian potential. This potential provides a sufficient energy barrier to prevent the condensate from reaching its thermodynamic ground state with zero angular momentum despite the presence of dissipation. Moreover, it is likely that vortex pinning by the Gaussian barrier is the physical mechanism responsible for the overall increase in the probability of observing vortices in the toroidal case as compared with the harmonic case. Additional toroidal trap statistics are provided in the Supplementary Information.

In the Supplementary Information, we provide movies of simulated condensate formation for quenches A and C. Here we describe one run in which a single vortex persists to the end of a quench A simulation in the harmonic trap. After the system temperature is initially lowered, the atomic density fluctuates temporally and spatially, as illustrated in Fig. 4a. A bulk BEC then begins to grow, and a tangle of vortices is trapped within the BEC as shown in Fig. 4b, in qualitative agreement with the models of superfluid turbulence[9,18–22] and the Kibble–Zurek mechanism[1–3,8]. The condensate now forms rapidly, but with clear vortex cores as shown in Fig. 4c. This state eventually damps to a single core as seen in Fig. 4d.

In examining the relationship between the BEC transition and the Kibble–Zurek mechanism, one would ideally study vortex formation with widely varying BEC growth rates in order to test the predictions of the scaling of the vortex density[1–3]. However, in our harmonic traps, we have only succeeded in increasing BEC growth rates by a

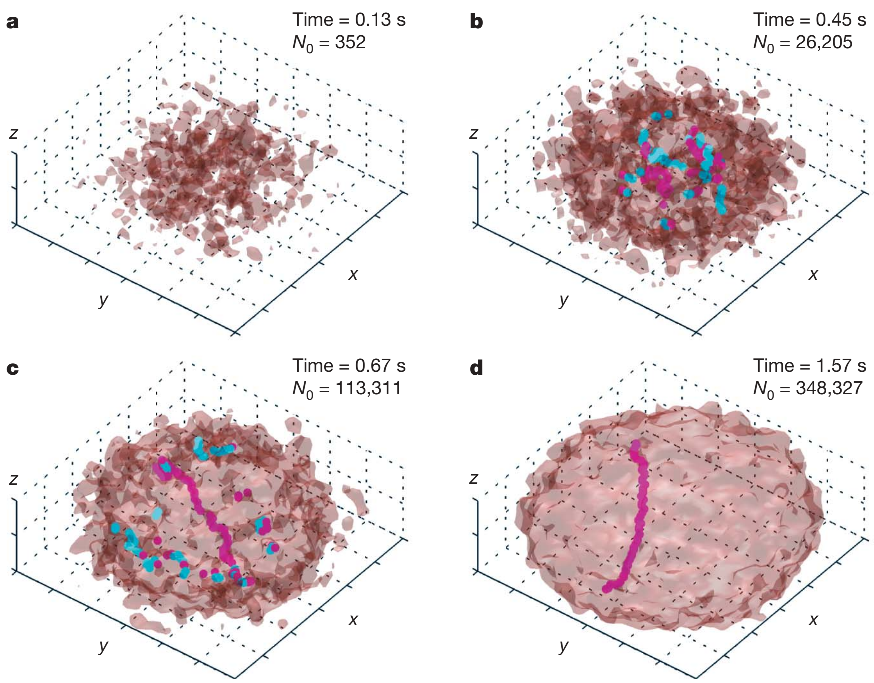


Figure 4 | BEC growth dynamics. **a–d**, Four snapshots during the simulated growth of a BEC showing isodensity surfaces (in light red) in a three-dimensional rendering. Vortex cores of opposite charges about the z axis are indicated as magenta and cyan lines. The corresponding times are 0.13 s (**a**), 0.45 s (**b**), 0.67 s (**c**), 1.57 s (**d**), where $t = 0$ s is the time when the quench is initiated in the simulation. The full movie from which these images were taken is provided as Supplementary Video 3.

factor of two to three, resulting in a factor of $\sim$2 increase in vortex formation compared with quenches A and B. In simulations, faster growth rates can potentially result in more vortices, as discussed in the Supplementary Information.

In addition to providing new experimental observations, our work places spontaneous topological defect formation on a theoretical foundation that has not been available in analogous studies in other systems. Related experimental investigations of spontaneous symmetry breaking of quenched ferromagnetic spinor BECs[29] may also yield insight into phase transition dynamics. The quantitative agreement between our experimental and theoretical results is of primary importance for their mutual interpretation: even in the ultra-cold BEC phase transition, thermal fluctuations can have an important role, and spontaneous topological defect formation may be virtually unavoidable in some situations. Our continuing work will explore in greater detail exactly how a condensate forms in this regime; the superfluid turbulence model describes vortex formation during condensation, but to what extent does the Kibble–Zurek mechanism's universality relate to the BEC transition? With further simulations and experiments, new details of the development of coherence in the birth of a superfluid may be uncovered, a tantalizing prospect addressing the interface between the classical and quantum worlds.

METHODS SUMMARY

In our experiments, ^{87}Rb atoms in the $|F=1, m_F=-1\rangle$ hyperfine state are confined in a time-averaged orbiting potential magnetic trap[30]. Evaporative cooling increases the phase space density to near the condensation critical point. The trap frequencies are then relaxed, and a final stage of cooling (quench A, B or C) induces the phase transition. In our numerical approach we describe the evolution of the condensate and its highly occupied excitations using a stochastic Gross–Pitaevski equation coupled to a thermal reservoir parameterized by a chemical potential μ and temperature T initially above the critical point. Evaporative cooling is simulated by a sudden change in μ and T. Further details regarding our approaches are provided in the full Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 March; accepted 12 August 2008.

1. Kibble, T. W. B. Topology of cosmic domains and strings. *J. Phys. A* **9,** 1387–1398 (1976).
2. Zurek, W. H. Cosmological experiments in superfluid helium? *Nature* **317,** 505–508 (1985).
3. Zurek, W. H. Cosmological experiments in condensed matter systems. *Phys. Rep.* **276,** 177–221 (1996).
4. Hendry, P. C., Lawson, N. S., Lee, R., McClintock, P. V. E. & Williams, C. D. H. Generation of defects in superfluid ^{4}He as an analogue of the formation of cosmic strings. *Nature* **368,** 315–317 (1994).
5. Ruutu, V. M. H. *et al.* Vortex formation in neutron-irradiated superfluid ^{3}He as an analogue of cosmological defect formation. *Nature* **382,** 334–336 (1996).
6. Baüerle, C., Bunkov, Y. M., Fisher, S. N., Godfrin, H. & Pickett, G. R. Laboratory simulation of cosmic string formation in the early Universe using superfluid ^{3}He. *Nature* **382,** 332–334 (1996).
7. Dodd, M., Hendry, P., Lawson, N., McClintock, P. & Williams, C. Nonappearance of vortices in the fast mechanical expansions of liquid ^{4}He through the Lambda transition. *Phys. Rev. Lett.* **81,** 3703–3706 (1998).
8. Anglin, J. R. & Zurek, W. H. Vortices in the wake of rapid Bose–Einstein condensation. *Phys. Rev. Lett.* **83,** 1707–1710 (1999).
9. Svistunov, B. V. Strongly non-equilibrium Bose–Einstein condensation in a trapped gas. *Phys. Lett. A* **287,** 169–174 (2001).
10. Stoof, H. T. C. Coherent versus incoherent dynamics during Bose–Einstein condensation in atomic gases. *J. Low Temp. Phys.* **114,** 11–108 (1999).
11. Davis, M. J., Ballagh, R. J. & Burnett, K. Dynamics of thermal Bose fields in the classical limit. *J. Phys. B* **34,** 4487–4512 (2001).
12. Davis, M. J., Morgan, S. A. & Burnett, K. Simulations of Bose fields at finite temperature. *Phys. Rev. Lett.* **87,** 160402 (2001).
13. Gardiner, C. W., Anglin, J. R. & Fudge, T. I. A. The stochastic Gross–Pitaevskii equation. *J. Phys. B* **35,** 1555–1582 (2002).
14. Gardiner, C. W. & Davis, M. J. The stochastic Gross–Pitaevskii equation: II. *J. Phys. B* **36,** 4731–4753 (2003).
15. Blakie, P. B. & Davis, M. J. Projected Gross–Pitaevskii equation for harmonically confined Bose gases at finite temperature. *Phys. Rev. A* **72,** 063608 (2005).
16. Davis, M. J. & Blakie, P. B. Critical temperature of a trapped Bose gas: Comparison of theory and experiment. *Phys. Rev. Lett.* **96,** 060404 (2006).
17. Bradley, A. S., Gardiner, C. W. & Davis, M. J. Bose–Einstein condensation from a rotating thermal cloud: Vortex nucleation and lattice formation. *Phys. Rev. A* **77,** 033616 (2008).
18. Svistunov, B. V. Highly nonequilibrium Bose condensation in a weakly interacting gas. *J. Mosc. Phys. Soc.* **1,** 373–390 (1991).
19. Kagan, Y., Svistunov, B. V. & Shlyapnikov, G. V. The Bose-condensation kinetics in an interacting Bose–gas. *Zh. Eksp. Teor. Fiz.* [*Sov. Phys. JETP* **75**, 387 (1992)] **101,** 528–539 (1992).
20. Kagan, Y. & Svistunov, B. V. Kinetics of long-range order formation in Bose-condensation in interacting gas. *Zh. Eksp. Teor. Fiz.* [*Sov. Phys. JETP* **78,** 187 (1994)] **105,** 353–367 (1994).
21. Kagan, Y. & Svistunov, B. V. Evolution of correlation properties and appearance of broken symmetry in the process of Bose–Einstein condensation. *Phys. Rev. Lett.* **79,** 3331–3334 (1997).
22. Berloff, N. G. & Svistunov, B. V. Scenario of strongly nonequilibrated Bose–Einstein condensation. *Phys. Rev. A* **66,** 013603 (2002).
23. Scherer, D. R., Weiler, C. N., Neely, T. W. & Anderson, B. P. Vortex formation by merging of multiple trapped Bose–Einstein condensates. *Phys. Rev. Lett.* **98,** 110402 (2007).
24. Matthews, M. R. *et al.* Vortices in a Bose–Einstein condensate. *Phys. Rev. Lett.* **83,** 2498–2501 (1999).
25. Madison, K. W., Chevy, F., Wohlleben, W. & Dalibard, J. Vortex formation in a stirred Bose–Einstein condensate. *Phys. Rev. Lett.* **84,** 806–809 (2000).
26. Davis, M. J. & Gardiner, C. W. Growth of a Bose–Einstein condensate: a detailed comparison of theory and experiment. *J. Phys. B* **35,** 733–742 (2002).
27. Pethick, C. & Smith, H. *Bose–Einstein Condensation in Dilute Gases* (Cambridge Univ. Press, 2002).
28. Ryu, C. *et al.* Observation of persistent flow of a Bose–Einstein condensate in a toroidal trap. *Phys. Rev. Lett.* **99,** 260401 (2007).
29. Sadler, L. E., Higbie, J. M., Leslie, S. R., Vengalattaore, M. & Stamper-Kurn, D. M. Spontaneous symmetry breaking in a quenched ferromagnetic spinor Bose–Einstein condensate. *Nature* **443,** 312–315 (2006).
30. Petrich, W., Anderson, M., Ensher, J. & Cornell, E. Stable, tightly confining magnetic trap for evaporative cooling of neutral atoms. *Phys. Rev. Lett.* **74,** 3352–3355 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Roberts, B. Svistunov, E. Wright and W. Zurek for discussions. The experimental work was financially supported by the US National Science Foundation under grant no. 0354977, and by the Army Research Office. The theoretical work was financially supported by the Australian Research Council Centre of Excellence for Quantum-Atom Optics and the University of Queensland.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to B.P.A. (bpa@optics.arizona.edu).

METHODS

Evaporative cooling. During the main evaporative cooling stages of our experimental procedure, our time-averaged orbiting potential (TOP) magnetic trap is created with a spherical quadrupole field that has a vertical magnetic field gradient of $B_z' = 300\,\mathrm{G\,cm^{-1}}$, and a magnetic bias field B_0 that has a direction that rotates in a horizontal plane at a frequency of either $\omega_{\mathrm{rot}} = 2\pi \times 4\,\mathrm{kHz}$ or $\omega_{\mathrm{rot}} = 2\pi \times 2\,\mathrm{kHz}$. Evaporative cooling proceeds over 72 s as B_0 decreases from 41 G to ~5.2 G, leaving a trapped cloud of atoms just above the condensation critical temperature T_c. The magnetic field gradient B_z' is then adiabatically reduced to $\sim 54\,\mathrm{G\,cm^{-1}}$ over 2 s, weakening the harmonic oscillator trapping frequencies to a measured radial (horizontal) trapping frequency of $\omega_r = 2\pi \times 7.8(\pm 0.1)$ Hz and an axial (vertical) trapping frequency of $\omega_z = 2\pi \times 15.3(\pm 0.2)$ Hz. The centre-of-mass position of the atom cloud correspondingly sags by about 0.6 mm vertically. In the final stage of our cooling cycle for quench A, we use a continuous 6-s ramp of the radiofrequency field, which evaporatively cools the atomic cloud from 70 nK to 20 nK, with $T_c \approx 42$ nK, to create condensates of $N_c \approx 5 \times 10^5$ atoms. For quench B, the continuous radiofrequency evaporative cooling ramp is replaced with a sudden jump to a final radiofrequency value, followed by a hold of the atomic sample in the trap before release and imaging. In this situation we find $T_c \approx 35$ nK and the final condensate number is $N_c \approx 3 \times 10^5$ atoms.

TOP trap. To ensure that the rotating bias field of the TOP trap plays no significant role in the spontaneous formation of vortices, we measured the z component of the net orbital angular momentum L_z of our condensates using surface wave spectroscopy. We excite a quadrupolar oscillation of the BEC in the horizontal plane, and stroboscopically probe the BEC with a set of non-destructive in-trap phase-contrast images[27], obtained by probing along the z axis[31,32]. The quadrupolar oscillations will then precess with a rate and direction proportional to L_z. In our measurements, there was no significant biasing of surface mode precession in a direction corresponding to the TOP trap rotation direction, an indication that TOP trap temporal dynamics have little to no influence on spontaneous vortex formation. This is discussed further in the Supplementary Information.

Toroidal trap. A potential-energy barrier was added to the centre of the magnetic trap using a focused blue-detuned laser beam with a wavelength of 660 nm, ~18 μW of power, and a ~6-μm Gaussian radius. The maximum beam intensity corresponds to a potential energy of roughly $k_B \times 20$ nK, where k_B is Boltzmann's constant. This can be compared with a chemical potential of about $k_B \times 10$ nK for our fully formed BECs in the purely harmonic trap. The beam was adiabatically ramped on before the final 6-s evaporation ramp, only slightly perturbing the thermal cloud but providing enough additional potential energy to exclude BEC atoms from the z axis of the trap.

Imaging. Our main imaging procedure involves the sudden removal of the magnetic trap, and the subsequent ballistic expansion of the trapped cloud. After 59 ms of expansion in the presence of an additional magnetic field to support the atoms against gravity, the atomic cloud is illuminated with near-resonant laser light, and the absorption profile of the atomic density distribution is imaged onto a camera. In our greyscale images, lighter shades represent higher optical depth, proportional to integrated column density along the line of sight in the z direction. A clear vortex core aligned along the z axis appears as a dark hole in the density distribution.

Stochastic Gross–Pitaevskii theory. We denote the condensate and low-energy portion of the trapped gas with the field $\alpha(\mathbf{x}, t)$, and define the Gross–Pitaevskii operator

$$L_{\mathrm{GP}} = -\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{x}) + g|\alpha(\mathbf{x}, t)|^2 \qquad (1)$$

where m is the mass of an atom, $V(\mathbf{x})$ is the trapping potential, $g = 4\pi\hbar^2 a/m$ characterizes the strength of atomic interactions, and a is the s-wave scattering length. The equation of motion for the field is

$$\mathrm{d}\alpha(\mathbf{x}, t) = \mathcal{P}\left\{ -\frac{i}{\hbar} L_{\mathrm{GP}}\alpha(\mathbf{x}, t)\mathrm{d}t + \frac{G(\mathbf{x})}{k_B T}(\mu - L_{\mathrm{GP}})\alpha(\mathbf{x}, t)\mathrm{d}t + \mathrm{d}W_G(\mathbf{x}, t)\right\} \qquad (2)$$

which has been derived from first principles using the Wigner phase-space representation[14]. The first term on the right describes unitary evolution of the classical field according to the Gross–Pitaevskii equation. The second term represents growth processes, that is, collisions that transfer atoms from the thermal bath to the classical field and vice versa, and the form of $G(\mathbf{x})$ may be determined from kinetic theory[17]. The third term is the complex-valued noise associated with condensate growth. The noise has Gaussian statistics and is defined by its only non-vanishing moment: $\langle \mathrm{d}W_G^*(\mathbf{x}, t)\mathrm{d}W_G(\mathbf{x}', t')\rangle = 2G(\mathbf{x})\mathrm{d}t\,\delta(\mathbf{x}-\mathbf{x}')\delta(t-t')$; it is also consistent with the fluctuation-dissipation theorem. The projection operator $\mathcal{P}$ restricts the dynamics to the low-energy region[12,15] defined by all harmonic oscillator modes with energy $\varepsilon < E_{\mathrm{cut}} = 40\hbar\omega_r$ for these calculations, which for our parameters gives about three particles per mode at the cut-off. For typical experimental parameters this method is accurate from slightly above the critical temperature to colder temperatures where there is still a significant thermal fraction[16].

The initial states used in our simulations are independent field configurations generated by ergodic evolution of the SGPE at equilibrium with the thermal cloud with $\mu_i = \hbar\omega_r$ and $T_i = 45\,(35)$ nK for quench A (B), representing the thermalized Bose gas above the transition temperature[16]. These parameters are then suddenly changed to final values chosen to match the final condensate number and temperature observed in the experiment: $\mu_f = 25\,(22)\,\hbar\omega_r$ and $T_f = 34\,(25)$ nK for quench A (B). We perform simulations for 300 (298) sets of initial conditions. By averaging over the different realizations we can calculate any quantum mechanical observable as a function of time, and in particular we diagonalize the single-particle density matrix to find the number of atoms in the condensate[15].

Because vortex formation is expected to depend on the BEC growth rate, which is difficult to calculate precisely, we adjust the coupling rate describing Bose-enhanced collisions between the classical field and thermal cloud to obtain a close match for the experimental BEC growth curves. We choose a spatially constant rate for the dimensionless coupling $\gamma = \hbar G(\mathbf{x})/k_B T$, shown to be a good approximation by Bradley *et al.*[17]. The noise at each time step then has the explicit form $\mathrm{d}W_G(\mathbf{x}, t) = \sum \phi_j(\mathbf{x})\sqrt{\gamma k_B T \mathrm{d}t/\hbar}\left(\eta_j + i\zeta_j\right)$, where $\phi_j(\mathbf{x})$ are the single-particle modes below the cut-off and the real Gaussian variables η_j, ζ_j are independent and have zero mean and unit variance. In principle γ is specified by a quantum Boltzmann integral, but here we treat it as an experimental fitting parameter for the condensate growth rate; it is never more than a factor of two different from the result of Eq. (A11) in Bradley *et al.*[17]. The effect of these parameter choices is discussed further in the Supplementary Information.

31. Chevy, F., Madison, K. W. & Dalibard, J. Measurement of the angular momentum of a rotating Bose–Einstein condensate. *Phys. Rev. Lett.* **85,** 2223–2227 (2000).
32. Haljan, P. C., Anderson, B. P., Coddington, I. & Cornell, E. A. Use of surface-wave spectroscopy to characterize tilt modes of a vortex in a Bose–Einstein condensate. *Phys. Rev. Lett.* **86,** 2922–2925 (2001).

LETTERS

Quantum oscillations in an overdoped high-T_c superconductor

B. Vignolle[1], A. Carrington[2], R. A. Cooper[2], M. M. J. French[2], A. P. Mackenzie[3], C. Jaudet[1], D. Vignolles[1], Cyril Proust[1] & N. E. Hussey[2]

The nature of the metallic phase in the high-transition-temperature (high-T_c) copper oxide superconductors, and its evolution with carrier concentration, has been a long-standing mystery[1]. A central question is how coherent electronic states, or quasiparticles, emerge from the antiferromagnetic insulator with doping. Recent quantum oscillation experiments on lightly doped copper oxides have shown evidence for small pockets of Fermi surface[2–5], the formation of which has been associated with the opening of the pseudogap—an anisotropic gap in the normal state excitation spectrum of unknown origin[1]. As the doping is increased, experiments suggest that the full Fermi surface is restored[6,7], although the doping level at which the pseudogap closes and the nature of the electronic ground state beyond this point have yet to be determined. Here we report the observation of quantum oscillations in the overdoped superconductor $Tl_2Ba_2CuO_{6+\delta}$ that show the existence of a large Fermi surface of well-defined quasiparticles covering two-thirds of the Brillouin zone. These measurements confirm that, in overdoped superconducting copper oxides, coherence is established at all Fermi wavevectors, even near the zone boundary where the pseudogap is maximal and electronic interactions are strongest; they also firmly establish the applicability of a generalized Fermi-liquid picture on the overdoped side of the superconducting phase diagram.

According to density functional theory (DFT) band structure calculations for hole-doped copper oxides, the electronic states within the CuO_2 planes form a large, almost cylindrical Fermi surface comprising $1 + p$ quasiparticles, where p is the number of doped holes per planar Cu ion[8,9]. In underdoped copper oxides, however, this band picture fails spectacularly owing to strong electron correlations and the opening of the pseudogap that depletes states in a manner yet to be fully understood[1]. Recently, small Fermi surface pockets were discovered in the underdoped yttrium-based copper oxides $YBa_2Cu_3O_{6.5}$ and $YBa_2Cu_4O_8$ through a series of quantum oscillation experiments[2–5]. In line with other bulk measurements, the area of these pockets was found to scale roughly with p, rather than $1 + p$, though neither their number nor their location are currently known. In marked contrast, angle-resolved photoemission spectroscopy (ARPES) measurements performed on other underdoped copper oxides suggest that the large underlying Fermi surface disintegrates into a set of disconnected Fermi arcs, centred on the zone diagonal, whose width diminishes with decreasing p and temperature T (ref. 10). This apparent disparity between the two experimental probes has yet to be resolved.

For heavily doped copper oxides, the situation is altogether different. Both interlayer angle-dependent magnetoresistance (ADMR)[6] and ARPES[7] measurements reveal a large quasi-two-dimensional Fermi surface in the single CuO_2 layer material $Tl_2Ba_2CuO_{6+\delta}$ (Tl2201), in broad agreement with DFT calculations[9] and earlier estimates from the low-temperature Hall coefficient[11]. Although ADMR is a useful probe of the underlying Fermi surface topology, it is a semi-classical effect and, as such, cannot give any definitive conclusions about the quantum mechanical nature of the electronic ground state. Similarly, although quasiparticle-like peaks have been seen in the ARPES-derived energy-density curves around the full Fermi surface of overdoped copper oxides[7,12], their widths near the zone boundaries are often too broad (up to 0.1 eV) to be considered as conclusive evidence for the long-lived fermionic quasiparticle states predicted by DFT.

From this standpoint, the observation of quantum oscillations is a much more powerful, definitive signature of coherent quasiparticles. Quantum oscillations are periodic oscillations in bulk physical properties—for example in the magnetoresistance (the Shubnikov–de Haas effect) or the magnetization (the de Haas–van Alphen (dHvA) effect)—that arise from the magnetic-field-induced quantization of the allowed energy levels of the quasiparticles. The quantization condition that underlies this effect requires the quasiparticle wavefunction to remain coherent throughout an entire orbit of the Fermi surface, and hence the observation of quantum oscillations is a crucial test of the fermionic nature of the low-lying electronic states.

The difficulty of observing quantum oscillations in a metal with a large Fermi surface should not be underestimated. The ADMR/ARPES measurements described above imply the existence of a Fermi surface sheet in Tl2201 with an average Fermi wavevector k_F a factor of six larger than that of the pocket first discovered in the yttrium-based copper oxides. The amplitude of the quantum oscillations is suppressed by impurity scattering as $\exp(-\pi\hbar k_F/eBl)$, where e is the electronic charge and l is the quasiparticle mean-free-path. At the combination of B and l (respectively 35 T and 165 Å) at which dHvA oscillations from the small Fermi surface first become visible in $YBa_2Cu_3O_{6.5}$ (ref. 5), the signal from a correspondingly large sheet would be damped by a further ten orders of magnitude. So for any experiment on overdoped copper oxides to be viable, it is essential to work at the highest magnetic fields, with optimized signal-to-noise ratio, on single crystals of the highest quality available.

The Tl2201 crystals used in this experiment (typical dimensions $0.2 \times 0.1 \times 0.02\ mm^3$) were grown via a self-flux method as described previously[13] and annealed in flowing oxygen at temperatures ranging from 270 °C to 450 °C for up one week in order to achieve a range of T_c values between 10 K and 30 K. Around 20 crystals were then screened for purity by measuring their zero-field interlayer resistivity $\rho_{\perp}(T)$ with a four-terminal a.c. method. Only four crystals with sufficiently large residual resistivity ratios (~20) were selected for the high-field magnetoresistance measurements

[1]Laboratoire National des Champs Magnétiques Pulsés (LNCMP), UMR CNRS-UPS-INSA 5147, Toulouse 31400, France. [2]H. H. Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, UK. [3]Scottish Universities Physics Alliance, School of Physics and Astronomy, University of St Andrews, St Andrews, Fife KY16 9SS, UK.

carried out in a standard ^{4}He cryostat at the LNCMP pulsed field facility in Toulouse using the same contact configuration. For the torque measurements, small pieces of the same crystals were attached to sensitive piezoresistive cantilevers and mounted in a dilution refrigerator in a second magnet cell at LNCMP.

Figure 1a, b shows interlayer resistance ($R_\perp$) data and magnetic torque (τ) data with the field oriented close to the c axis for two different Tl2201 crystals at temperatures below their zero-field superconducting transitions ($T_c \approx 10$ K). For both sets of data, an expanded view near the field maxima reveals clear oscillations whose amplitude grows with increasing field strength. Figure 1c shows the oscillatory part of the magnetization $\Delta\tau/B$ plotted versus inverse magnetic field $1/B$. The observation of oscillations, periodic in $1/B$, in both the magnetization and the resistivity, at fields well above the upper critical field B_{c2}, confirm these as quantum oscillations. Strikingly, as shown in Fig. 1c, these oscillations are more than one order of magnitude faster than those found in underdoped $YBa_2Cu_4O_8$.

The frequency (F) of the oscillations is directly related to the extremal cross-sectional area A of the Fermi surface normal to the field orientation, via the Onsager relation, $A = 2\pi eF/\hbar$. Figure 2a shows the fast Fourier transform of the data in Fig. 1c. A single, sharp dHvA frequency of 18,100 ± 50 T is obtained, corresponding to a Fermi surface extremal cross-sectional area A of $172.8 \pm 0.5\ \mathrm{nm}^{-2}$ and average k_F of $7.42 \pm 0.05\ \mathrm{nm}^{-1}$. The temperature dependence of the Shubnikov–de Haas amplitude is shown in Fig. 2b. Fitting this with the standard Lifshitz–Kosevich theory, we obtain a cyclotron effective mass $m^* = 4.1 \pm 1.0\ m_e$, where m_e is the free electron mass. Finally, Fig. 2c shows that the field dependence of the amplitude of the dHvA oscillations follows the expected exponential decay, from which we estimate $l = 320$ Å.

All these numbers are in excellent agreement with those deduced from other measurements in the same material with similar doping levels. The Fermi surface topology deduced from ADMR[6,14] is reproduced as a solid black line in Fig. 2a inset. Its area agrees extremely well with the measured dHvA frequency, and corresponds to ~65% of the Brillouin zone of Tl2201 and a doping level $p = 0.30$. In addition, given that for a two-dimensional Fermi surface, the electronic specific heat (Sommerfeld coefficient) is $\gamma_{el} = (\pi N_A k_B^2 a^2/3\hbar^2)m^*$ (where k_B is the Boltzmann constant, N_A is Avogadro's number, and $a = 3.86$ Å is the in-plane lattice constant[15]), our value of m^* corresponds to $\gamma_{el} \approx 6.0 \pm 1.0\ \mathrm{mJ\ mol^{-1}\ K^{-2}}$, in excellent agreement with that measured directly[16] for overdoped polycrystalline Tl2201 ($7 \pm 2\ \mathrm{mJ\ mol^{-1}\ K^{-2}}$). Hence, in contrast to the previous work[2–5], we can make quantitative comparisons between quasiparticle properties derived from quantum oscillations at high fields and those measured directly by transport and thermodynamics at zero field. This good overall consistency suggests that the Fermi surface is composed entirely of the single quasi-two-dimensional sheet that we observe. DFT calculations predict that the bare band mass in stoichiometric Tl2201 is $\sim 1.2 m_e$ (ref. 9). The difference between the measured and calculated masses implies strong electron-correlation-driven renormalization, even at this elevated doping level.

Despite strong electron–electron interactions, the observation of quantum oscillations implies that quasiparticles exist at all points on the Fermi surface of overdoped Tl2201. The observation of genuine quantum oscillations in Tl2201 supports the recognized idea that generalized Fermi-liquid theory can be applied on the overdoped side of the phase diagram. Moreover, the observation of quantum oscillations on both sides of optimal doping, albeit with very different frequencies, suggests that the quantum oscillations observed in underdoped copper oxides directly probe the Fermi surface there, rather than some anomalous vortex physics[17], and raises the intriguing prospect that quasiparticle states may survive, at least at some loci on the Fermi surface, across the entire doping range, from the insulating/superconducting boundary to the non-superconducting metal on the heavily overdoped side.

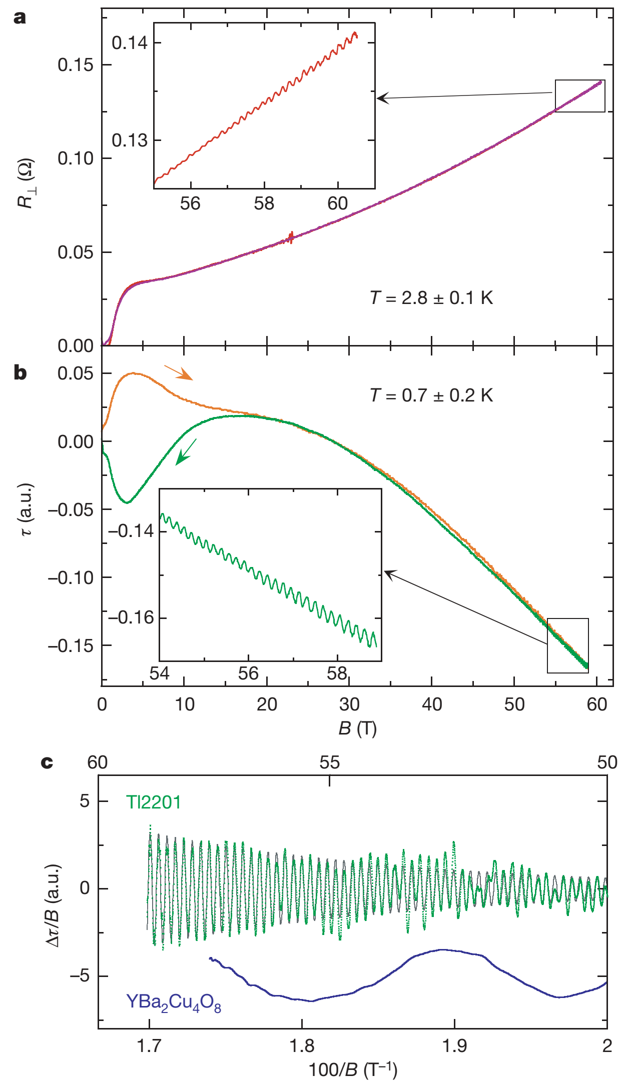


Figure 1 | Quantum oscillations in Tl2201. **a**, Raw data on interlayer resistance ($R_\perp$) with **B**//c for a Tl2201 single crystal at $T = 2.8 \pm 0.1$ K. $R_\perp(B)$ rises rapidly above the irreversibility field B_{irr}, passes through a small plateau, then grows quasi-quadratically with field up to 60 T. Inset, magnified view of the high-field region of the down sweep. Small but well-defined oscillations are clearly resolved in the raw data, with an amplitude that grows with increasing field strength. The maximum amplitude of the oscillations is only 0.5 mΩ. **b**, Averaged magnetic torque data (from five sweeps at temperatures between 0.6 K and 0.8 K) with **B** close to the c axis for a different Tl2201 crystal. The temperatures of the torque sweeps are subject to an additional uncertainty of 150 mK due to the weak thermal link inside the dilution refrigerator[5] and the high currents needed to observe the oscillations. Below $B = 14$ T, the torque shows hysteretic behaviour due to flux trapping and expulsion in the superconducting mixed state. Again, well-defined oscillations are clearly resolved in the expanded region shown in the inset. The value of T_c for both crystals is 10 K (defined by their zero-resistive state), compared with the maximal T_c in Tl2201 of 92 K. The torque crystal showed a very small kink in the zero-field $\rho_\perp$ data around 20 K, suggesting that some fraction of the crystal, presumably the surface layer, had a higher T_c value. Note that the difference in B_{irr} exhibited by both crystals is amplified in the pulsed magnetic field because the sweep rate dB/dt for the higher B_{irr} sample is greater. a.u., arbitrary units. **c**, The oscillatory component of the torque data shown in **b** plotted as green dots against $1/B$ (the corresponding B values are shown at the top of the panel), after a monotonic background has been subtracted. The black line superimposed on the data is a fit to the Lifshitz–Kosevich expression for dHvA oscillations in a two-dimensional metal[29]. Also shown (blue), for comparison, is the oscillatory component in the torque signal on underdoped $YBa_2Cu_4O_8$.

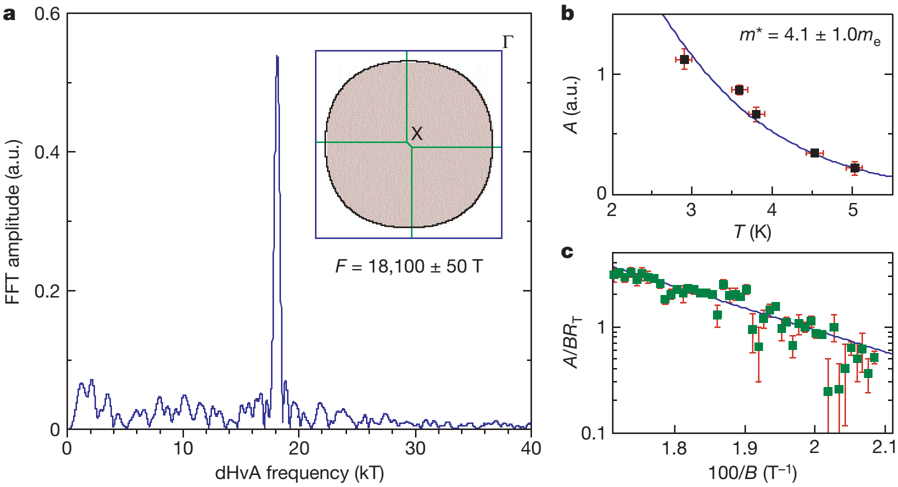


Figure 2 | Features of the oscillation data. **a**, Fast Fourier transform (FFT) of the data shown in Fig. 1c (field range 50–58.8 T), revealing a single, sharp dHvA frequency of $F = 18{,}100 \pm 50$ T, corresponding to an average Fermi surface radius $k_F = 7.42 \pm 0.05\,\mathrm{nm}^{-1}$. Inset, a cross-section of the Fermi surface topology of overdoped Tl2201 ($T_c = 15$ K) deduced from ADMR[6,14]. The area of this tubular Fermi surface is in excellent agreement with our measured dHvA frequency. **b**, Temperature dependence of the Shubnikov–de Haas amplitude. According to the standard Lifshitz–Kosevich expression for the oscillatory magnetization[29], the thermal damping factor $R_T = X/\sinh(X)$, where $X = (2\pi^2 k_B/\hbar e)m^* T/B$ and m^* is the quasiparticle effective mass that is enhanced over the band-mass by many-body interactions. Error bars, 1σ. **c**, Field dependence of the amplitude of the dHvA oscillations shown in Fig. 1c, divided by R_T. Each point represents a fit of 1.5 oscillations to $A/(R_T B)\sin(2\pi F/B + \phi)$. From the fit to the exponential decay, we tentatively estimate a mean-free-path of $l = 320$ Å. The actual mean-free-path may be up to a factor of two longer than this, because the limited field range of our measurements does not allow us to rule out a low frequency beat with another close frequency, as expected for a quasi-two-dimensional Fermi surface such as this[30]. The transport mean-free-path in the best crystals, as estimated from in-plane resistivity measurements[11], is of the order $l = 670$ Å. Error bars are deduced from a combination of systematic error and standard deviation.

Another noteworthy feature of this result is that quantum oscillations have been observed in a metal for which one key signature of a Landau Fermi-liquid, namely a purely quadratic temperature dependence of the electrical resistivity at low temperatures, is absent. According to ADMR experiments on overdoped Tl2201 crystals with comparable T_c values, the scattering rate contains two temperature-dependent components; a dominant isotropic component, which varies quadratically with temperature and is characteristic of fermionic quasiparticle scattering, and an anisotropic component, maximal near the Brillouin zone boundaries and varying linearly with temperature down to very low temperatures[14], consistent with the observed form of the in-plane resistivity[11,18]. This unusual form of the scattering rate has been associated with proximity to a quantum critical point[19,20] or to the Mott insulating state[21], but its origin is as yet unknown. The fact that quasiparticles arise, despite there being a linear-in-T contribution to the scattering rate, is in agreement with theoretical predictions[22,23]. However, to actually see the effect of the anomalous contribution to the self energy in the temperature dependence of the oscillations would require following them to much lower temperature[23,24].

Finally, our measurements offer strong support for the scenario[25] that beyond a critical doping level within the superconducting dome, the pseudogap vanishes—that is, that the pseudogap and the superconducting gap are not coincident in the overdoped regime. We stress here that closure of the pseudogap is not field-induced, as the Fermi surface parameters we derive are entirely consistent with zero-field transport[11], thermodynamic[17] and spectroscopic[7] data. The task now is to determine whether this large Fermi surface evolves into a collection of small pockets or into a series of disconnected arcs at low doping. In other words, does the carrier density decrease smoothly from $1 + p$ to p, or does a competing order—such as antiferromagnetism[26], d-density-wave[27], orbital loops[20] or stripes[28]—cause a Fermi surface reconstruction along some critical line in the (p, T) phase diagram? Given the excellent agreement between dHvA and ARPES results evident in overdoped Tl2201, combined measurements on the same underdoped compound seem to be essential to help resolve this long-standing controversy.

Received 4 June; accepted 28 July 2008.

1. Timusk, T. & Statt, B. The pseudogap in high temperature superconductors: An experimental survey. *Rep. Prog. Phys.* **62**, 61–122 (2000).
2. Doiron-Leyraud, N. *et al.* Quantum oscillations and the Fermi surface in an underdoped high-T_c superconductor. *Nature* **447**, 565–568 (2007).
3. Yelland, E. A. *et al.* Quantum oscillations in the underdoped cuprate $YBa_2Cu_4O_8$. *Phys. Rev. Lett.* **100**, 047003 (2008).
4. Bangura, A. F. *et al.* Small Fermi surface pockets in underdoped high temperature superconductors: Observation of Shubnikov-de Haas oscillations in $YBa_2Cu_4O_8$. *Phys. Rev. Lett.* **100**, 047004 (2008).
5. Jaudet, C. *et al.* de Haas–van Alphen oscillations in the underdoped high-temperature superconductor $YBa_2Cu_3O_{6.5}$. *Phys. Rev. Lett.* **100**, 187005 (2008).
6. Hussey, N. E. *et al.* A coherent three-dimensional Fermi surface in a high transition temperature superconductor. *Nature* **425**, 814–817 (2003).
7. Platé, M. *et al.* Fermi surface and quasiparticle excitations of overdoped $Tl_2Ba_2CuO_{6+\delta}$. *Phys. Rev. Lett.* **95**, 077001 (2005).
8. Andersen, O. K., Liechtenstein, A. I., Jepsen, O. & Paulsen, F. LDA energy bands, low energy Hamiltonians, t', t'', $t_\perp$(**k**) and $J_\perp$. *J. Phys. Chem. Solids* **56**, 1573–1591 (1995).
9. Singh, D. J. & Pickett, W. E. Electronic characteristics of $Tl_2Ba_2CuO_6$. *Physica C* **203**, 193–199 (1992).
10. Kanigel, A. *et al.* Evolution of the pseudogap from Fermi arcs to the nodal metal. *Nature Phys.* **2**, 447 (2006).
11. Mackenzie, A. P., Julian, S. R., Sinclair, D. C. & Lin, C. T. Normal state magnetotransport in superconducting $Tl_2Ba_2CuO_{6+\delta}$ to millikelvin temperatures. *Phys. Rev. B* **53**, 5848–5854 (1996).
12. Yusof, Z. M. *et al.* Quasiparticle liquid in the highly overdoped $Bi_2Sr_2CaCu_2O_{8+\delta}$. *Phys. Rev. Lett.* **88**, 167006 (2002).
13. Tyler, A. W. *An Investigation into the Magnetotransport Properties of Layered Superconducting Perovskites.* Ph.D. thesis, Univ. Cambridge (1997).
14. Abdel-Jawad, M. *et al.* Anisotropic scattering and anomalous normal state transport in a high temperature superconductor. *Nature Phys.* **2**, 821–825 (2006).
15. Mackenzie, A. P. *et al.* Calculation of thermodynamic and transport properties of Sr_2RuO_4 at low temperatures using known Fermi surface parameters. *Physica C* **263**, 510–515 (1996).
16. Loram, J. W. *et al.* The electronic specific heat of cuprate superconductors. *Physica C* **235–240**, 134–137 (1994).
17. Alexandrov, A. S. Theory of quantum magneto-oscillations in underdoped cuprate superconductors. *J. Phys. Condens. Matter* **20**, 192202 (2008).
18. Proust, C., Boaknin, E., Hill, R. W., Taillefer, L. & Mackenzie, A. P. Heat transport in a strongly overdoped cuprate: Fermi liquid and a pure *d*-wave BCS superconductor. *Phys. Rev. Lett.* **89**, 147003 (2002).
19. Dell'Anna, L. & Metzner, W. Electrical resistivity near Pomeranchuk instability in two dimensions. *Phys. Rev. Lett.* **98**, 136402 (2007).

20. Zhu, L., Aji, V., Shekhter, A. & Varma, C. M. Universality of single-particle spectra of cuprate superconductors. *Phys. Rev. Lett.* **100,** 057001 (2008).
21. Ossadnik, M., Honerkamp, C., Rice, T. M. & Sigrist, M. Breakdown of Landau theory in overdoped cuprates near the onset of superconductivity. Preprint at ⟨http://arXiv.org/abs/0805.3489⟩ (2008).
22. Wasserman, A., Springford, M. & Han, F. The de Haas-van Alphen effect in a marginal Fermi liquid. *J. Phys. Condens. Matter* **3,** 5335–5339 (1991).
23. Wasserman, A. & Springford, M. The influence of many-body interactions on the de Haas-van Alphen effect. *Adv. Phys.* **45,** 471–503 (1996).
24. McCollam, A., Julian, S. R., Rourke, P. M. C., Aoki, D. & Flouquet, J. Anomalous de Haas-van Alphen oscillations in $CeCoIn_5$. *Phys. Rev. Lett.* **94,** 186401 (2005).
25. Tallon, J. L. & Loram, J. W. The doping dependence of T^* — what is the true high-T_c phase diagram? *Physica C* **349,** 53–68 (2001).
26. Lin, J. & Millis, A. J. Theory of low-temperature Hall effect in electron-doped cuprates. *Phys. Rev. B* **72,** 214506 (2005).
27. Chakravarty, S. *et al.* Hidden order in the cuprates. *Phys. Rev. B* **63,** 094503 (2001).
28. Millis, A. J. & Norman, M. R. Antiphase stripe order as the origin of electron pockets observed in 1/8-hole-doped cuprates. *Phys. Rev. B* **76,** 220503 (2007).
29. Schoenberg, D. *Magnetic Oscillations in Metals* (Cambridge Univ. Press, 1984).
30. Bergemann, C., Mackenzie, A. P., Julian, S. R., Forsythe, D. & Ohmichi, E. Quasi-two-dimensional Fermi-liquid properties of the unconventional superconductor Sr_2RuO_4. *Adv. Phys.* **52,** 639–725 (2003).

Acknowledgements We acknowledge technical and scientific assistance from A. Audouard, M. Nardone, N. Shannon and D. C. Sinclair. This work was supported by EPSRC, LNCMP, the French ANR IceNET and EuroMagNET.

Author Contributions The crystals were grown by A.P.M. and screened by M.M.J.F. and N.E.H. The experiments were carried out by B.V., A.C., R.A.C., C.P., C.J., D.V. and N.E.H., and the analysis performed by B.V., A.C., M.M.J.F., C.P. and N.E.H.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to N.E.H. (n.e.hussey@bristol.ac.uk).

Vol 455|16 October 2008|doi:10.1038/nature07320

LETTERS

Bottom-up organic integrated circuits

Edsger C. P. Smits[1,2,3], Simon G. J. Mathijssen[2,4], Paul A. van Hal[2], Sepas Setayesh[2], Thomas C. T. Geuns[2], Kees A. H. A. Mutsaers[2], Eugenio Cantatore[5], Harry J. Wondergem[2], Oliver Werzer[6], Roland Resel[6], Martijn Kemerink[4], Stephan Kirchmeyer[7], Aziz M. Muzafarov[8], Sergei A. Ponomarenko[8], Bert de Boer[1], Paul W. M. Blom[1] & Dago M. de Leeuw[1,2]

Self-assembly—the autonomous organization of components into patterns and structures[1]—is a promising technology for the mass production of organic electronics. Making integrated circuits using a bottom-up approach involving self-assembling molecules was proposed[2] in the 1970s. The basic building block of such an integrated circuit is the self-assembled-monolayer field-effect transistor (SAMFET), where the semiconductor is a monolayer spontaneously formed on the gate dielectric. In the SAMFETs fabricated so far, current modulation has only been observed in submicrometre channels[3–5], the lack of efficient charge transport in longer channels being due to defects and the limited intermolecular π–π coupling between the molecules in the self-assembled monolayers. Low field-effect carrier mobility, low yield and poor reproducibility have prohibited the realization of bottom-up integrated circuits. Here we demonstrate SAMFETs with long-range intermolecular π–π coupling in the monolayer. We achieve dense packing by using liquid-crystalline molecules consisting of a π-conjugated mesogenic core separated by a long aliphatic chain from a monofunctionalized anchor group. The resulting SAMFETs exhibit a bulk-like carrier mobility, large current modulation and high reproducibility. As a first step towards functional circuits, we combine the SAMFETs into logic gates as inverters; the small parameter spread then allows us to combine the inverters into ring oscillators. We demonstrate real logic functionality by constructing a 15-bit code generator in which hundreds of SAMFETs are addressed simultaneously. Bridging the gap between discrete monolayer transistors and functional self-assembled integrated circuits puts bottom-up electronics in a new perspective.

The fabrication of SAMFETs has been attempted by several groups. Systems investigated were functionalized acenes on aluminium oxide[4], hexabenzocoronene on silicon dioxide[3], and oligothiophenes with a functionalized short aliphatic linker on aluminium oxide and silicon dioxide[5]. Electrical transport was measured as a function of channel length. No current was measured when using long channels. Current modulation due to the field effect was only observed in submicrometre channels. The accumulation layer could not be pinched off, and there was limited current saturation. Moreover, the mobility and the yields were low and the reproducibility poor. The lack of efficient charge transport was due to the presence of defects and to the limited intermolecular π−π coupling between the molecules in the self-assembled monolayers. A prerequisite for functional SAMFETs is a dense and ordered semiconducting monolayer.

In a SAMFET the monolayer is formed on the gate dielectric. We use atomically flat amorphous silicon dioxide. Usually trichlorosilanes or trialkoxysilanes are used as anchoring groups. A self-assembled monolayer (SAM) can be formed by a condensation reaction with hydroxyl groups on the hydrolysed silicon dioxide surface. Defects are formed because of uncontrolled self-condensation, prohibiting long-range order[6]. To prevent these defects, monofunctional anchoring groups are crucial. Dimers formed upon self-condensation do not interfere with the SAM formation on the gate dielectric.

We use α-substituted quinquethiophene as the core of the semiconducting molecule, for several reasons. First, these oligothiophenes have charge carrier mobilities several orders of magnitude higher than the corresponding β-substituted molecules[7]. Second, the mobility increases with the number of thiophene units[8]. However, the solubility then significantly decreases, hampering processability. As a result, quinquethiophene is a good compromise between solubility and charge carrier mobility. The semiconducting core is (α, ω)-functionalized with aliphatic chains. This spacer concept originates from liquid crystals, from which it is well known that the aliphatic chain helps anisotropic ordering of rod-like mesogenic groups[9–11]. We use an undecane spacer between the thiophene core and the anchoring group. The conformational degree of freedom allows the molecule to self-assemble and optimize its π−π stacking. For stability and solubility reasons, we attach an ethane chain to the other side. The chemical structure is presented in the inset of Fig. 1a. The synthetic details and characterization, as well as the full data set of molecules investigated, are presented in the Supplementary Information. There we discuss the fact that the semiconductor is a mixture of the functionalized compound and an inactive impurity that cannot bind to the silicon dioxide gate dielectric. Co-crystallization is unlikely but cannot yet be excluded by chemical analysis.

We fabricate discrete SAMFETs on heavily doped silicon wafers, acting as a common gate, with 200 nm of thermally grown silicon dioxide, and construct gold source and drain contacts using conventional photolithographic methods. Titanium is used as an adhesion layer. The necessary use of the titanium layer can inhibit injection. In the Supplementary Information we show, in focused-ion-beam transmission electron spectroscopy images (FIB-TEM), that the gold electrodes are under-etched. The titanium adhesion layer at the edge of the electrode is dissolved. The electrode is collapsed and the gold makes close contact with the silicon dioxide. Despite the use of the titanium adhesion layer, charge injection occurs through the gold contact.

We activate the silicon dioxide surface using an oxygen plasma treatment followed by acid hydrolysis, and form the SAM by submerging the substrate into a dry toluene solution of the semiconducting molecule. After the monolayer formation, the substrate is thoroughly rinsed and dried.

[1]Molecular Electronics, Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands. [2]Philips Research Laboratories, High Tech Campus 4, 5656 AE Eindhoven, The Netherlands. [3]Dutch Polymer Institute, PO Box 902, 5600 AX Eindhoven, The Netherlands. [4]Department of Applied Physics, [5]Mixed-Signal Microelectronics Group, Department of Electrical Engineering, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands. [6]Institute of Solid State Physics, Graz University of Technology, Petersgasse 16A, 8010 Graz, Austria. [7]H. C. Starck GmbH, Chemiepark Leverkusen, Building B202, 51368 Leverkusen, Germany. [8]Enikolopov Institute of Synthetic Polymer Materials, Russian Academy of Sciences, Profsoyuznaya 70, 117393 Moscow, Russia.

Atomic force microscopy (AFM) measurements show a smooth continuous layer. The layer thickness is determined to be about 3 nm using X-ray photoelectron spectroscopy. To probe the electron density profile in the SAM, we performed X-ray reflectivity measurements. The reflected X-ray intensity is presented in Fig. 1a as a function of wavevector transfer, q_z. The high-frequency oscillations are interference patterns due to the 200-nm silicon dioxide layer. The longer-wavelength oscillations with maxima at about 0.14 and 0.30 Å^{-1} originate in the SAM layer. We characterize the SAM by comparing the measured data with a calculated reflectivity curve. The SAM is modelled as a bilayer with two different electron densities. The bottom layer corresponds to the aliphatic chain and the top layer to the thiophene core. Figure 1a shows that a good agreement is obtained. The thicknesses of the two regions are fitted to be 15.6 and 20.6 Å, respectively. The numbers obtained coincide with the calculated lengths of the undecane aliphatic spacer and the quinquethiophene backbone, respectively. The reflectivity measurements demonstrate a perfect out-of-plane order with the molecules standing almost perpendicular to the substrate.

We determine the in-plane order from grazing-incidence diffraction measurements. The diffracted intensity as a function of the in-plane scattering vector, q_{xy}, is presented in Fig. 1b. The inset shows the diffracted intensity perpendicular to the substrate, q_z, as a function of q_{xy}. The horizontal line at low q_z is due to diffuse scattering from the sample surface at the critical angle as first described by Yoneda[12]. The vertical lines are the so-called Bragg rods[13], showing in-plane order in the SAM layer. The rods are observed at scattering vectors that are indexed as the (1.1), (0.2) and (1.2) reflections of a rectangular unit cell with lattice constants of 5.49 and 7.69 Å. The peak indices and unit-cell dimensions are in agreement with a unit cell that contains two molecules packed in a herringbone structure, as is commonly observed for oligothiophenes[14]. The occurrence of Bragg rods is due to the absence of periodicity perpendicular to the ordered layer. This indicates that our SAM layer is a dense, smooth monolayer.

Having established the order in the SAM, we characterize the electrical transport properties. Before the measurements, the SAMFETs are annealed in a dynamic vacuum at 120 °C for one hour to remove residual water and solvents. A typical transfer characteristic for a ring transistor with a channel length of 40 μm and a channel width of 1,000 μm is presented in Fig. 2a. The *p*-type SAMFET shows

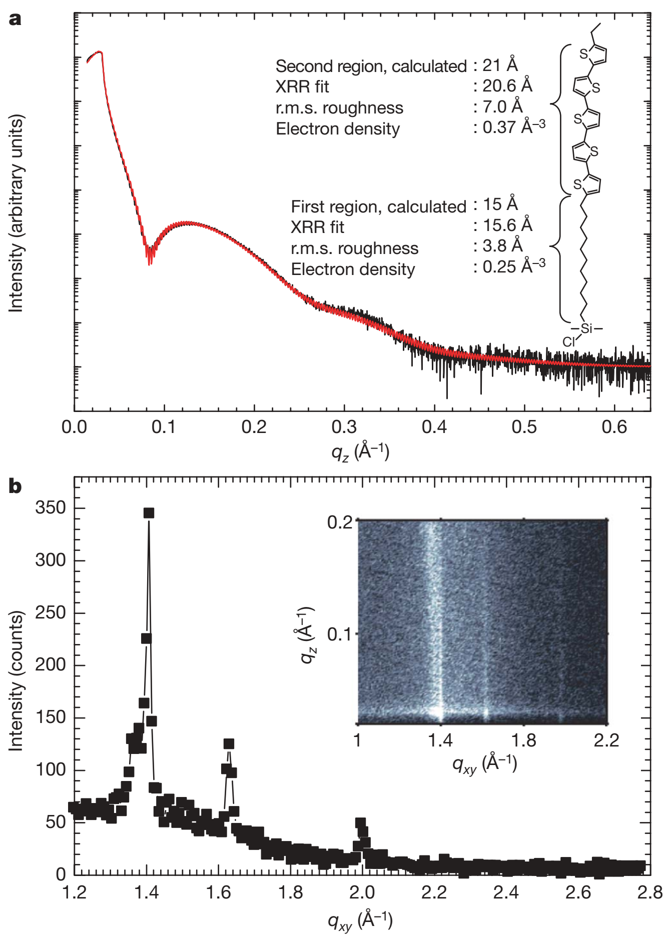


Figure 1 | SAM microstructure. **a**, X-ray reflectivity (XRR) of the SAM on silicon dioxide. The solid red line is the fit to the experimental data. The inset shows the chemical structure of the molecule and the thicknesses, root-mean-square (r.m.s.) roughness values and electron densities obtained from the fit. The measurements reveal a layer thickness equivalent to one monolayer. **b**, Synchrotron grazing-incidence diffraction measurements showing the diffracted intensity as a function of in-plane scattering vector, q_{xy}. The inset shows the presence of Bragg rods at in-plane scattering vectors of 1.407, 1.635 and 1.997 Å^{-1}, indicative of two-dimensional crystalline in-plane order in the SAM.

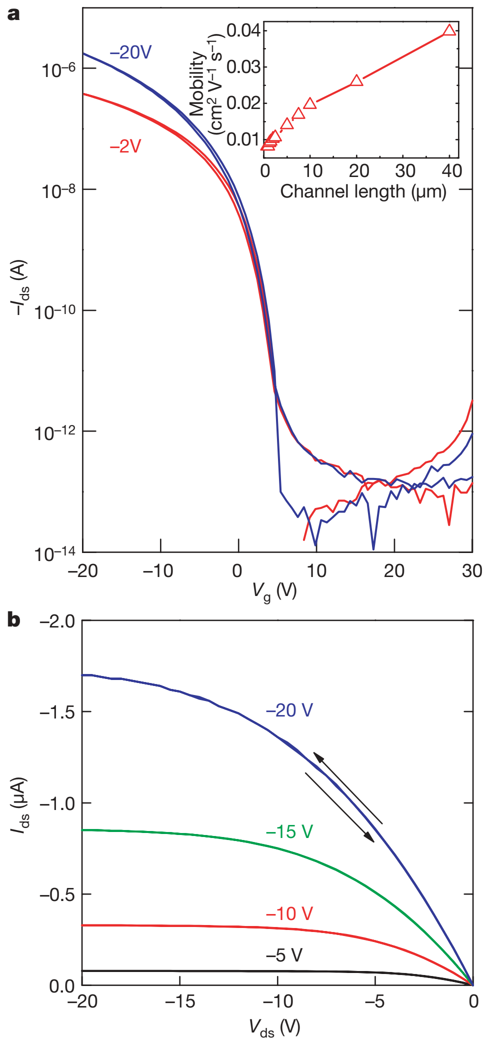


Figure 2 | Electrical SAMFET transport characteristics. **a**, Linear and saturated transfer characteristics of a SAMFET with a channel length of 40 μm and a channel width of 1,000 μm, using drain biases of −2 and −20 V, respectively. I_{ds}, drain–source current; V_g, gate voltage. The inset shows the linear mobility as a function of channel length. The mobility increases from 0.01 cm^2 V^{-1} s^{-1} for 0.75-μm channels to 0.04 cm^2 V^{-1} s^{-1} at 40-μm channels. **b**, Output characteristics for the corresponding SAMFET. The gate voltage was varied from −5 to −20 V in steps of −5 V. V_{ds}, drain–source voltage.

almost no hysteresis. The linear and saturated mobility is 0.04 cm^2 V^{-1} s^{-1}. The mobility is thermally activated with an activation energy of about 80 meV. The mobility is comparable to that obtained from quinquethiophene single-crystalline thin-film transistors[8,15,16]. This is significant, considering that the charge transport in the SAMFET occurs through a single layer a few nanometres thick. The transfer curve shows a current modulation of seven decades. This large on–off current modulation is a special feature of SAMFETs: it is a result of the absence of bulk conductivity in a monolayer. The use of long channels and the negligible off currents yield saturated output curves with high output resistance as shown in Fig 2b.

The inset of Fig. 2a shows the mobility as a function of channel length. The mobility increases with channel length to 0.04 cm^2 V^{-1} s^{-1} for a 40-μm channel. In homogeneous semiconductors, the mobility does not depend on channel length. In actual transistors, the mobility increases with channel length owing to the presence of contact resistance. In an inhomogeneous semiconductor, where the charge transport is dominated by defects such as islands and grain boundaries, the mobility decreases with channel length. The scaling in the inset of Fig. 2a shows that the transport is not defect limited but instead is homogeneous through the SAM.

We also prepared SAMFETs using short immersion times and lower concentrations of the molecule. The surface coverage was estimated from AFM measurements. The mobility decreases when the SAM is not a dense layer. Furthermore, SAMFETs with incomplete monolayers show a mobility that decreases with channel length. This scaling behaviour is expected in a two-dimensional percolation problem[16–18]. Hence, to obtain a high-mobility SAMFET it is crucial to have a uniform, dense monolayer.

We extracted the transport parameters of numerous SAMFETs. The yield is about unity. Our statistical analysis shows reproducible values for key device parameters such as mobility, threshold voltage, subthreshold slope and on–off current modulation (see Supplementary Information). Typically a standard deviation in the mobility of only 0.005 cm^2 V^{-1} s^{-1} is found.

The small parameter spread allows the combination of SAMFETs into integrated circuits. Fabrication of logic gates and integrated circuits requires a patterned gate and vertical interconnects. To this end we developed a 150-mm process technology on the basis of a doped polysilicon gate, gold electrodes, a silicon dioxide gate dielectric and photolithographically defined interconnects. It is well known that surfaces of polysilicon overgrown with thermal oxide are rough. However, we find, counterintuitively, that the mobility of the SAMFET is remarkably insensitive to the interface roughness. Circuits were fabricated by self-assembling the semiconductor onto the substrate.

Figure 2a shows that the SAMFETs are normally on at 0-V gate bias. Therefore, all logic gates used to build the digital circuits presented here are based on so-called unipolar $V_{gs} = 0$ inverters where the load transistor has the source connected to the gate[19]. A circuit diagram and the input–output characteristics are presented in Fig. 3. The inverters show voltage amplification with a gain that increases with the d.c. supply voltage. The noise margin is about 1 V. The values are similar to those of state-of-the-art unipolar thin-film organic inverters[20]. The small parameter spread in mobility and threshold voltage allow us to combine SAMFET inverters into seven-stage ring oscillators. We measure a switching frequency of 5 kHz at a supply voltage of −10 V. Real logic functionality is demonstrated with the realization of 15-bit code generators. The integrated circuits combine over 300 SAMFETs, and contain an onboard clock generator, hard-wired memory, a four-bit counter, decoder logic and a load modulator. The output of a 15-bit code generator is presented in Fig. 4. The bit rate is around 1 kbit s^{-1} at a supply voltage of −40 V. The circuit performance is similar to that of state-of-the-art organic integrated circuits developed for organic radio-frequency identification transponders[20].

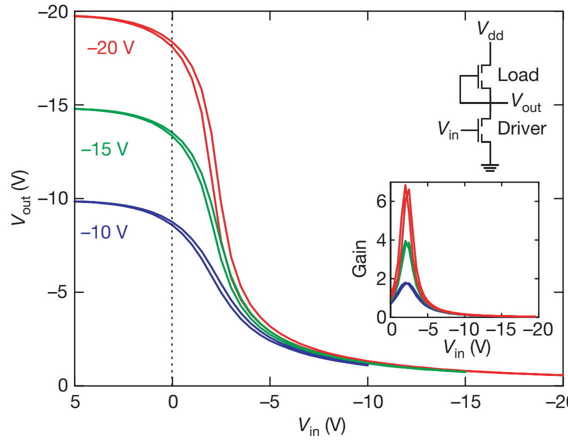


Figure 3 | SAMFET inverter. Static input–output characteristics of a SAMFET-based unipolar $V_{gs} = 0$ inverter. V_{in}, input voltage; V_{out}, output voltage. The inverter was measured with supply voltages, V_{dd}, of −10, −15 and −20 V. The inset shows a diagram of the logic gate and a plot of the measured gain (colour coded as for main panel).

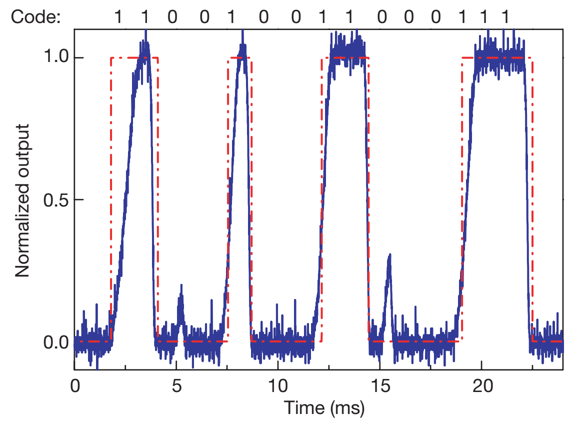


Figure 4 | Integrated circuit. Output of a 15-bit code generator based on SAMFETs. The bit rate is about 1 kbit s^{-1} at a supply voltage of −40 V. The outputted code is indicated at the top and by the red line.

In summary, we have constructed the basic building block of bottom-up organic electronics, the SAMFET, demonstrating high mobility and on–off current modulation. The excellent reproducibility allows the production of functional integrated circuits in which hundreds of SAMFETs are addressed simultaneously.

METHODS SUMMARY

Full details on methods are presented in the Supplementary Information. The molecular design is discussed, a historical perspective is given and the full set of synthesized and investigated molecules is included.

We fabricated discrete SAMFETs on atomically smooth silicon dioxide. The formation of a single monolayer was confirmed by X-ray photoemission spectroscopy, AFM investigations on fully and partially covered monolayers, and by X-ray reflectivity measurements. The in-plane order of the SAM was studied using grazing-incidence diffraction. The occurrence of Bragg peaks unambiguously proves that the SAM layer is a dense, smooth monolayer. Scanning Kelvin probe microscopy measurements on partially covered SAMFETs show that only the molecules connected to the electrodes participate in the electrical transport. We demonstrate using FIB-TEM images that charge injection into the SAM occurs through the gold contact.

Scaling of the mobility with channel length is presented and the contact resistance is estimated by transmission line modelling. Partially covered SAMFETs show inverse scaling. Integrated circuits have been made using polysilicon gate electrodes overgrown with thermal oxide of which the surface roughness is addressed. We show that the mobility of the SAMFET is remarkably insensitive to the interface roughness. Finally, a section on circuit design modelling and layout is included, and measures to minimize parasitic currents are presented.

Received 1 February; accepted 7 August 2008.

1. Whitesides, G. M. & Grzybowski, B. Self-assembly at all scales. *Science* **295,** 2418–2421 (2002).
2. Aviram, A. & Ratner, M. A. Molecular rectifiers. *Chem. Phys. Lett.* **29,** 277–283 (1974).
3. Guo, X. *et al.* Chemoresponsive monolayer transistors. *Proc. Natl Acad. Sci. USA* **103,** 11452–11456 (2006).
4. Tulevski, G. S. *et al.* Attaching organic semiconductors to gate oxides: in situ assembly of monolayer field effect transistors. *J. Am. Chem. Soc.* **126,** 15048–15050 (2004).
5. Mottaghi, M. *et al.* Low-operating-voltage organic transistors made of bifunctional self-assembled monolayers. *Adv. Funct. Mater.* **17,** 597–604 (2007).
6. Onclin, S., Ravoo, B. J. & Reinhoudt, D. N. Engineering silicon oxide surfaces using self-assembled monolayers. *Angew. Chem. Int. Edn* **44,** 6282–6304 (2005).
7. Garnier, F. *et al.* Molecular engineering of organic semiconductors: design of self-assembly properties in conjugated thiophene oligomers. *J. Am. Chem. Soc.* **115,** 8716–8721 (1993).
8. Halik, M. *et al.* relationship between molecular structure and electrical performance of oligothiophene organic thin film transistors. *Adv. Mater.* **15,** 917–922 (2003).
9. van Breemen, A. J. J. M. *et al.* Large area liquid crystal monodomain field-effect transistors. *J. Am. Chem. Soc.* **128,** 2336–2345 (2006).
10. Ponomarenko, S. A. *et al.* Star-shaped oligothiophenes for solution-processible organic electronics: flexible aliphatic spacers approach. *Chem. Mater.* **18,** 4101–4108 (2006).
11. McCulloch, I. *et al.* Liquid-crystalline semiconducting polymers with high charge-carrier mobility. *Nature Mater.* **5,** 328–333 (2006).
12. Yoneda, Y. Anomalous surface reflection of X rays. *Phys. Rev.* **131,** 2010–2013 (1963).
13. Fenter, P. in *Self-Assembled Monolayers of Thiols* (ed. Uhlman, A.) 111–147 (Academic, 1991).
14. Fichou, D. *Handbook of Oligo- and Polythiophenes* (Wiley-VCH, 1999).
15. Melucci, M. *et al.* Multiscale self-organization of the organic semiconductor α-quinquethiophene. *J. Am. Chem. Soc.* **125,** 10266–10274 (2003).
16. Dinelli, F. *et al.* Spatially correlated charge transport in organic thin film transistors. *Phys. Rev. Lett.* **92,** 116802 (2004).
17. Ruiz, R., Papadimitratos, A., Mayer, A. C. & Malliaras, G. G. Thickness dependence of mobility in pentacene thin-film transistors. *Adv. Mater.* **17,** 1795–1798 (2005).
18. Park, B.-N. Seo, S. & Evans, P. G. Channel formation in single-monolayer pentacene thin film transistors. *J. Phys. D* **40,** 3506–3511 (2007).
19. E. Cantatore & E. J. & Meijer. in *ESSCIRC '03* 29–36 (Proc. 29th Eur. Solid State Circuits Conf., IEEE, 2003).
20. Cantatore, E. *et al.* A 13.56-MHz RFID system based on organic transponders. *IEEE J. Solid State Circuits.* **42,** 84–92 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge H. Nulens and C. van der Marel for the AFM and X-ray photoemission spectroscopy measurements, M. Kaiser for the FIB-TEM analysis, A. P. Pleshkova for the mass spectrometry analysis, F. Zontone for technical assistance and N. Willard for discussions. We acknowledge financial support from the Dutch Polymer Institute, project 516, the EU project NAIMO (NMP4-CT-2004-500355), the Dutch Technology Foundation STW, the Austrian Science Foundation and H. C. Starck GmbH. We thank the European Synchrotron Research Facility for the use of the beamline ID10B.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence should be addressed to D.M.d.L. (dago.de.leeuw@philips.com). Requests for materials should be addressed to S.K. (stephan.kirchmeyer@hcstarck.com).

LETTERS

Oxidation state of iron in komatiitic melt inclusions indicates hot Archaean mantle

Andrew J. Berry[1], Leonid V. Danyushevsky[2], Hugh St C. O'Neill[3], Matt Newville[4] & Stephen R. Sutton[4,5]

Komatiites are volcanic rocks mainly of Archaean age that formed by unusually high degrees of melting of mantle peridotite. Their origin is controversial and has been attributed to either anhydrous melting of anomalously hot mantle[1–3] or hydrous melting at temperatures only modestly greater than those found today[4,5]. Here we determine the original $Fe^{3+}/\Sigma Fe$ ratio of 2.7-Gyr-old komatiitic magma from Belingwe, Zimbabwe[6], preserved as melt inclusions in olivine, to be 0.10 ± 0.02, using iron K-edge X-ray absorption near-edge structure spectroscopy. This value is consistent with near-anhydrous melting of a source with a similar oxidation state to the source of present-day mid-ocean-ridge basalt. Furthermore, this low $Fe^{3+}/\Sigma Fe$ value, together with a water content of only 0.2–0.3 wt% (ref. 7), excludes the possibility that the trapped melt contained significantly more water that was subsequently lost from the inclusions by reduction to H_2 and diffusion. Loss of only 1.5 wt% water by this mechanism would have resulted in complete oxidation of iron (that is, the $Fe^{3+}/\Sigma Fe$ ratio would be ~1). There is also no petrographic evidence for the loss of molecular water. Our results support the identification of the Belingwe komatiite as a product of high mantle temperatures (~1,700 °C), rather than melting under hydrous conditions (3–5-wt% water), confirming the existence of anomalously hot mantle in the Archaean era.

The oxidation state of a magma is given by the ratio of Fe^{3+} to Fe^{2+} (conveniently expressed as $Fe^{3+}/\Sigma Fe$), and reflects the nature of the source and partial melting conditions[8]. $Fe^{3+}/\Sigma Fe$ affects melt structure, viscosity and the temperature and composition of crystallizing phases. It also affects the speciation and amount of volatiles degassed during volcanic eruptions—particularly sulphur-bearing gases such as SO_2—and, thus, volcanogenic forcing of global climate[9]. However, the final oxidation state measured in an igneous rock reflects not only that of the parental magma but also the influence of many subsequent processes, including degassing at the Earth's surface, alteration and metamorphism. Inclusions of melt trapped in early-crystallizing minerals such as olivine potentially preserve nearly pristine, pre-eruptive, undegassed samples of the magma, but the typically small size of such inclusions (<50 μm) has hitherto precluded the precise determination of $Fe^{3+}/\Sigma Fe$. Here we accurately obtain $Fe^{3+}/\Sigma Fe$ from individual melt inclusions.

$Fe^{3+}/\Sigma Fe$ in silicate glasses can be determined by redox titrations, commonly referred to as 'wet chemistry'[10,11], and Mössbauer spectroscopy[12], but these methods lack spatial resolution and are generally restricted to bulk samples. Electron energy-loss spectroscopy[13] and the electron microprobe[14] have the spatial resolution desired, but the techniques are unlikely to be suitable for the analysis of glass inclusions, owing to the high electron beam fluxes involved and the likelihood of beam-induced changes in oxidation state occurring. X-ray absorption near-edge structure (XANES) spectroscopy, which has micrometre spatial resolution, is non-destructive, element specific and requires no sample preparation other than a polished surface, has promised much for a number of years[15], but accurate quantification of $Fe^{3+}/\Sigma Fe$ has been hindered by the need for compositionally matched standards and a lack of sufficient precision in the limited $Fe^{3+}/\Sigma Fe$ range of most terrestrial samples[15–17]. However, it has been shown recently that oxidation state ratios in glasses can be determined with great accuracy and precision, by preparing a series of standards as a function of oxygen fugacity (f_{O_2}) that allows $M^{n+}/\Sigma M$ (for an element M with oxidation state $n+$) to be systematically varied in a way that is generally not possible for minerals, owing to the constraints of crystal chemistry[18,19].

The preparation of glass standards of mid-ocean-ridge basalt (MORB), tholeiite, pantellerite and andesite compositions, and the acquisition of XANES spectra are described in the Methods. XANES spectra for the MORB standard glasses are shown in Fig. 1a. The spectra comprise an absorption edge and a pre-edge feature corresponding to the $1s \rightarrow 3d$ transition. This feature, which is shown in detail in Fig. 1b, comprises transitions to the crystal field levels of Fe^{2+} and Fe^{3+} in the various coordination environments present in the glass[20]. It is possible to reproduce the pre-edge by quantitatively modelling these transitions, using a line shape predicted from the instrumental (Gaussian) and electronic (Lorentzian) contributions to the spectral resolution, but this is unnecessary for quantifying the $Fe^{3+}/\Sigma Fe$ ratio. Instead, as seen in Fig. 1b, the feature shifts to higher energy with increasing oxidation state, owing to the change in energy of the final crystal field levels, allowing an empirical calibration curve to be constructed; the transition intensity changes with the site symmetry[18,21,22].

The key to determining the pre-edge energy precisely is deconvolution from the absorption-edge baseline. Here the pre-edge was fit to two pseudo-Voigt peaks, constrained to have the same width and line shape, simultaneously with the baseline, which was modelled as the tail of two other pseudo-Voigt functions. An example fit is shown in Fig. 1c. The intensity-weighted energy, or centroid, of the baseline-subtracted pre-edge can then be correlated with the $Fe^{3+}/\Sigma Fe$ value of the standards. The resulting calibration curve, obtained using all the compositions studied, is shown in Fig. 2. The functional form of the calibration may vary with the method used to model the baseline; what is important for accurate and precise $Fe^{3+}/\Sigma Fe$ determinations is that the standards and unknown contain iron at similar sites (which is assumed for similar compositions cooled at similar rates) and that all spectra are processed identically.

The validity of the approach was tested by determining $Fe^{3+}/\Sigma Fe$ for a number of olivine-hosted MORB melt inclusions. The inclusions were rapidly heated (to 1,250 °C in 5 min) and quenched to produce a glass. The similar heating and quenching paths of both the unknown and the standard should produce similar iron coordination environments, thus allowing quantification of the unknown from the

[1]Department of Earth Science and Engineering, Imperial College London, South Kensington, SW7 2AZ, UK. [2]CODES, University of Tasmania, Hobart, Tasmania 7001, Australia. [3]Research School of Earth Sciences, Australian National University, Canberra, ACT 0200, Australia. [4]Center for Advanced Radiation Sources, [5]Department of Geophysical Sciences, University of Chicago, Chicago, Illinois 60637, USA.

standard calibration curve. The average $Fe^{3+}/\Sigma Fe$ for the MORB inclusions using the calibration shown in Fig. 2 was 0.14 ± 0.03, in excellent agreement with the value of 0.13 ± 0.02 determined by wet chemical analysis in ref. 11, where it was also found that the f_{O_2} of MORB does not vary between oceanic settings or enriched or depleted melts. This constant value of $Fe^{3+}/\Sigma Fe$ was reproduced for melt inclusions by our XANES method.

The methodology was then applied to determining the $Fe^{3+}/\Sigma Fe$ ratio of komatiite melt inclusions. Komatiites are rare, mantle-derived melts (restricted to the Archaean, with a few exceptions) that are characterized by MgO contents >18 wt%, and require a very high degree of partial mantle melting (up to 50%) in comparison with, for example, MORB (10–20%). Two models have been proposed to explain their origin: the plume, or 'hot-melting', model, involving decompression melting of a mantle source up to 500 °C hotter than today[1–3]; and the 'wet-melting' model, in which komatiites are Archaean analogues of modern subduction-related magmas with a mantle temperature only 100 °C hotter than today[4,5].

It should be possible to distinguish between the two models if well-preserved products of original komatiite melt were available. Unfortunately, most komatiites have been extensively altered as a result of their age. However, unusually fresh komatiite lava flows have been found within the Archaean Belingwe belt, Zimbabwe, which contain olivine phenocrysts ($Fo_{89.0–93.5}$) that have preserved inclusions of the original melt[6,7,23]. We heated the inclusions, which at room temperature have recrystallized into an aggregate of olivine, pyroxene and interstitial glass, to 1,300 °C in 30 s and quenched them to give a homogeneous glass (Fig. 3). The composition of the glass differs from that of the originally trapped melt, owing to crystallization of olivine on the inclusion walls that is not re-dissolved on heating[24,25] and post-entrapment equilibration with the host olivine[7,25]. Both of these effects will shift the $Fe^{3+}/\Sigma Fe$ ratio of the glass to higher values, and the values determined here thus represent upper limits. The inclusions were prepared as free-standing sections (which are also ideal for the determination of water contents by infrared spectroscopy).

The pre-edge region of the XANES spectrum of a komatiite melt inclusion is shown in Fig. 1b. It is apparent, in relation to MORB, that the inclusion is reduced. Using the calibration curve in Fig. 2, the average $Fe^{3+}/\Sigma Fe$ value of four komatiite inclusions was determined to be 0.10 ± 0.02 (the individual values were 0.10 ± 0.01, 0.11 ± 0.03, 0.08 ± 0.01 and 0.10 ± 0.01).

Fertile mantle peridotite contains 0.3 wt% Fe_2O_3, implying a $Fe^{3+}/\Sigma Fe$ value of 0.035 ± 0.005 for a source with 6.3 wt% ΣFe (refs 26, 27). Assuming bulk melt–residue partition coefficients of 1 for FeO and 0.1 for Fe_2O_3 (ref. 26), 15% batch melting of peridotite, which is appropriate for the genesis of primitive MORB, should result in a melt with $Fe^{3+}/\Sigma Fe \approx 0.13$. This is in agreement with what is routinely observed, demonstrating that MORBs reliably reflect the oxidation state of their sources. Higher degrees of partial melting lower the $Fe^{3+}/\Sigma Fe$ value of the magma by dilution; with the same partition coefficients, the $Fe^{3+}/\Sigma Fe$ range of 0.08–0.12 found here for the komatiite melt inclusions would correspond to 16–33% partial melting. Even greater degrees of partial melting are inferred if, as is probable, the partition coefficient of Fe_2O_3 decreases with increasing melting as pyroxenes, the main crystalline hosts of this component, are progressively eliminated from the residue[28]. The oxidation state of the Belingwe komatiite source is therefore, within uncertainty, the same as the source of modern MORB, in agreement with that deduced from V/Sc systematics[29]. By contrast, arc peridotites indicate that hydrated

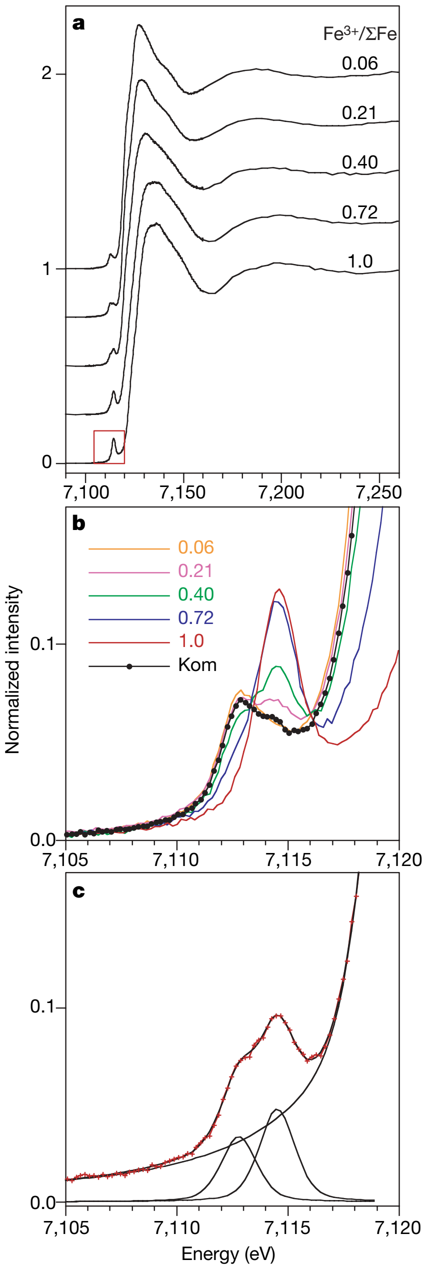


Figure 1 | Iron K-edge XANES spectra of quenched silicate melts. a, Spectra of synthetic MORB glass standards with the $Fe^{3+}/\Sigma Fe$ values indicated. Spectra have been offset for clarity. **b**, XANES pre-edge region, indicated by the box in **a**, showing the $1s \rightarrow 3d$ transition of both the MORB standards and a natural komatiite melt inclusion (Kom). **c**, Fit to the pre-edge of the MORB standard with $Fe^{3+}/\Sigma Fe = 0.40$, showing the data (symbols), the baseline and the pseudo-Voigt components.

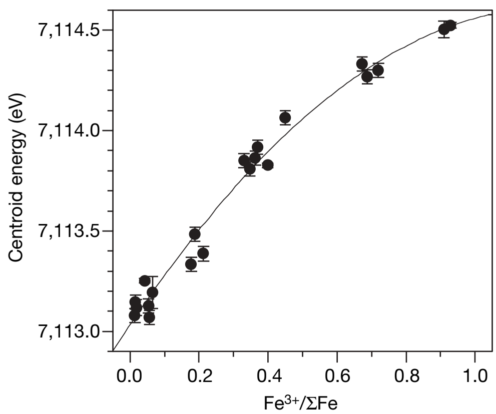


Figure 2 | Calibration curve for determining $Fe^{3+}/\Sigma Fe$ from the XANES pre-edge centroid energy. The curve is derived from a second-order polynomial fit to data for MORB, tholeiite, pantellerite and andesite standards; error bars, 1 s.d., determined from replicate measurements.

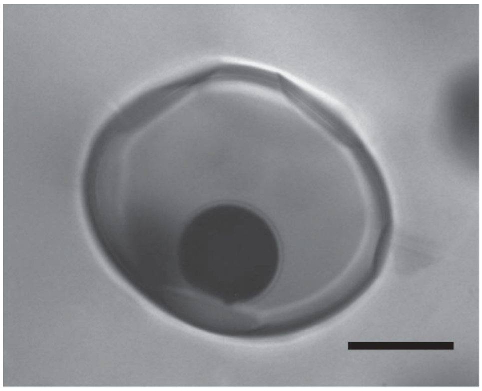


Figure 3 | Optical image of an olivine-hosted melt inclusion (after heating) from a komatiite of the Belingwe belt, Zimbabwe. The inclusion comprises quenched melt (glass) and a dark 'shrinkage bubble'; scale bar, 20 μm.

mantle is much more oxidized[30], leading to the characteristic highly oxidized signature of subduction-related magmas[8].

The measured water contents of the komatiite inclusions range from 0.18 to 0.26 wt% (ref. 7), which is far less than the 3 to 5 wt% required by the wet-melting model. However, melt inclusions can change their water contents after entrapment, through loss by diffusion of H_2 or water (ref. 24). Loss of water results in the formation of haloes of tiny fluid inclusions and recrystallized host mineral rims[31]. No such structures were observed in the komatiite samples studied. The most probable process, however, involves reduction of water to H_2, which may then diffuse rapidly out of the sample. Such reduction must result in the concomitant oxidation of another species, with iron being the only polyvalent element present in sufficient abundance to form the other half of the redox pair; that is, $H_2O_{(melt)} + 2FeO_{(melt)} = H_{2(gas)} + Fe_2O_{3(melt)}$. Loss of water by H_2 diffusion is therefore associated with an increase in $Fe^{3+}/\Sigma Fe$, and complete oxidation of an inclusion with 12 wt% FeO would be expected (that is, $Fe^{3+}/\Sigma Fe = 1$) for the loss of only 1.5 wt% water.

The komatiite melt inclusions therefore show no evidence for the loss of water either directly or by reduction and diffusion of H_2, and the measured water contents are likely to be those of the original melt. These results are consistent with the 'anomalously hot mantle' model for the formation of the Belingwe komatiite. The results also show that despite the higher temperature of the komatiite source, its oxidation state was similar to that of present-day mantle, which rules out significant interactions with either the core (more reduced) or crustal material (more oxidized); it also argues against increases in the oxidation state of the mantle over geological time, contrary to some hypotheses (ref. 32 and references therein). The XANES technique is ideally suited to determining the $Fe^{3+}/\Sigma Fe$ ratio of melt inclusions, as it provides the necessary spatial resolution, the samples are isotropic and, hence, orientation independent, and compositionally matched standards can be prepared using the methods of experimental petrology. The method opens up the possibility for the routine determination of redox state in geological samples on the same analytical scale as is currently achievable for major elements by the electron microprobe and trace elements by ion microprobe and laser-ablation mass spectrometry.

METHODS SUMMARY

Synthetic glass standards were prepared by equilibrating reagent-grade oxide mixes in a 1-atm gas mixing furnace at 1,400 °C, using $CO/CO_2/O_2$ to control f_{O_2} at log values relative to the nickel–nickel-oxide buffer (NNO) of −5, −2, +1, +3 and +6, followed by quenching in water[18,19]. Samples were also prepared at NNO + 11 by equilibrating with PtO_2 in a sealed platinum capsule at 10 kbar and 1,400 °C using a piston–cylinder apparatus. The $Fe^{3+}/\Sigma Fe$ ratio of each glass can be estimated using an empirical expression that relates Fe^{3+}/Fe^{2+} to f_{O_2}, temperature, pressure and composition[33]. In this way, standards of any melt composition with any $Fe^{3+}/\Sigma Fe$ ratio can be obtained. Glasses with a pantellerite composition were prepared by equilibrating natural material in gas-mixing experiments; the $Fe^{3+}/\Sigma Fe$ ratios of these samples were determined by redox titration. All standards were thinned to produce doubly polished, free-standing sections with a thickness of 50–100 μm, which is close to the 1/e absorption path length of iron K-edge X-rays.

Iron K-edge XANES spectra were recorded at beamline 13-ID-C, GeoSoilEnviroCARS (GSECARS, University of Chicago), of the Advanced Photon Source (Argonne National Laboratory, USA). Owing to the large penetration depth of the analysis beam, it was necessary to consider the entire beam trajectory through the sample and the potential for excitation of the host olivine ($Mg_{1.8}Fe_{0.2}SiO_4$). This was avoided by preparing inclusions such that they are exposed on both sides of the section and orienting the sample normal to the beam, which was focused to a spot with a diameter of ~5 μm.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 March; accepted 26 August 2008.

1. Green, D. H. Genesis of Archean peridotitic magmas and constraints on Archean geothermal gradients and tectonics. *Geology* **3,** 15–18 (1975).
2. Nisbet, E. G., Cheadle, M. J., Arndt, N. T. & Bickle, M. J. Constraining the potential temperature of the Archean mantle - a review of the evidence from komatiites. *Lithos* **30,** 291–307 (1993).
3. Arndt, N. *et al.* Were komatiites wet? *Geology* **26,** 739–742 (1998).
4. Allegre, C. J. in *Komatiites* (eds Arndt, N. T. & Nisbet, E. G.) 495–500 (Springer, 1982).
5. Parman, S. W., Grove, T. L. & Dann, J. C. The production of Barberton komatiites in an Archean subduction zone. *Geophys. Res. Lett.* **28,** 2513–2516 (2001).
6. Nisbet, E. G. *et al.* Uniquely fresh 2.7 Ga komatiites from the Belingwe greenstone-belt, Zimbabwe. *Geology* **15,** 1147–1150 (1987).
7. Danyushevsky, L. V., Gee, M. A. M., Nisbet, E. G. & Cheadle, M. J. Olivine-hosted melt inclusions in Belingwe komatiites: implications for cooling history, parental magma composition and its H_2O content. *Geochim. Cosmochim. Acta* **66** (Suppl. 1), A168 (2002).
8. Carmichael, I. S. E. The redox states of basic and silicic magmas: a reflection of their source region. *Contrib. Mineral. Petrol.* **106,** 129–141 (1991).
9. Self, S., Thordarson, T. & Widdowson, M. Gas fluxes from flood basalt eruptions. *Elements* **1,** 283–287 (2005).
10. Christie, D. M., Carmichael, I. S. E. & Langmuir, C. H. Oxidation states of mid-ocean ridge basalt glasses. *Earth Planet. Sci. Lett.* **79,** 397–411 (1986).
11. Bezos, A. & Humler, E. The $Fe^{3+}/\Sigma Fe$ ratios of MORB glasses and their implications for mantle melting. *Geochim. Cosmochim. Acta* **69,** 711–725 (2005).
12. Jayasuriya, K. D., O'Neill, H. St C., Berry, A. J. & Campbell, S. J. A Mössbauer study of the oxidation state of iron in silicate melts. *Am. Mineral.* **89,** 1597–1609 (2004).
13. Garvie, L. A. J. & Buseck, P. R. Ratios of ferrous to ferric iron from nanometre-sized areas in minerals. *Nature* **396,** 667–670 (1998).
14. Fialin, M., Bézos, A., Wagner, C., Magnien, V. & Humler, E. Quantitative electron microprobe analysis of $Fe^{3+}/\Sigma Fe$: basic concepts and experimental protocol for glasses. *Am. Mineral.* **89,** 654–662 (2004).
15. Delaney, J. S., Dyar, M. D., Sutton, S. R. & Bajt, S. Redox ratios with relevant resolution: solving an old problem by using the synchrotron microXANES probe. *Geology* **26,** 139–142 (1998).
16. Dyar, M. D., Delaney, J. S. & Sutton, S. R. Fe XANES spectra of iron-rich micas. *Eur. J. Mineral.* **13,** 1079–1098 (2001).
17. Bonnin-Mosbah, M. *et al.* Iron oxidation states in silicate glass fragments and glass inclusions with a XANES micro-probe. *J. Non-Cryst. Solids* **288,** 103–113 (2001).
18. Berry, A. J., O'Neill, H. St C., Jayasuriya, K. D., Campbell, S. J. & Foran, G. J. XANES calibrations for the oxidation state of iron in a silicate glass. *Am. Mineral.* **88,** 967–977 (2003).
19. Berry, A. J. & O'Neill, H. St C. A XANES determination of the oxidation state of chromium in silicate glasses. *Am. Mineral.* **89,** 790–798 (2004).
20. Westre, T. E. *et al.* A multiplet analysis of Fe K-edge 1s→3d pre-edge features of iron complexes. *J. Am. Chem. Soc.* **119,** 6297–6314 (1997).
21. Wilke, M., Farges, F., Petit, P.-E., Brown, G. E. Jr & Martin, F. Oxidation state and coordination of Fe in minerals: an Fe *K*-XANES spectroscopic study. *Am. Mineral.* **86,** 714–730 (2001).
22. Wilke, M., Partzsch, G. M., Bernhardt, R. & Lattar, D. Determination of the iron oxidation state in basaltic glasses using XANES at the K-edge. *Chem. Geol.* **213,** 71–87 (2004).
23. McDonough, W. F. & Ireland, T. R. The intraplate origin of komatiites inferred from trace elements in glass inclusions. *Nature* **365,** 432–434 (1993).
24. Danyushevsky, L. V., McNeill, A. W. & Sobolev, A. V. Experimental and petrological studies of melt inclusions in phenocrysts from mantle-derived magmas: an overview of techniques, advantages and complications. *Chem. Geol.* **183,** 5–24 (2002).
25. Danyushevsky, L. V., Sokolov, S. & Falloon, T. J. Melt inclusions in olivine phenocrysts: using diffusive re-equilibration to determine the cooling history of a crystal, with implications for the origin of olivine-phyric volcanic rocks. *J. Petrol.* **43,** 1651–1671 (2002).

26. Canil, D. *et al.* Ferric iron in peridotites and mantle oxidation states. *Earth Planet. Sci. Lett.* **123,** 205–220 (1994).
27. Palme, H. & O'Neill, H. St C. in *The Mantle and Core* (ed. Carlson, R. W.) 1–38 (Treatise on Geochemistry 2, Elsevier, 2003).
28. Canil, D. & O'Neill, H. St C. Distribution of ferric iron in some upper-mantle assemblages. *J. Petrol.* **37,** 609–635 (1996).
29. Canil, D. Vanadium partitioning and the oxidation state of Archaean komatiite magmas. *Nature* **389,** 842–845 (1997).
30. Parkinson, I. J. & Arculus, R. J. The redox state of subduction zones: insights from arc-peridotites. *Chem. Geol.* **160,** 409–423 (1999).
31. Danyushevsky, L. V., Berry, A. J., O'Neill, H. St C., Newville, M. & Sutton, S. R. Fe^{3+}/Fe^{2+} of melt inclusions: implications for melt H_2O contents. *Geochim. Cosmochim. Acta* **71,** A200 (2007).
32. Galimov, E. M. Redox evolution of the Earth caused by a multistage formation of its core. *Earth Planet. Sci. Lett.* **233,** 263–276 (2005).
33. Kress, V. C. & Carmichael, I. S. E. The compressibility of silicate liquids containing Fe_2O_3 and the effect of composition, temperature, oxygen fugacity and pressure on their redox states. *Contrib. Mineral. Petrol.* **108,** 82–92 (1991).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. R. Scott for sample preparation, N. Métrich for providing the pantellerite glass standards and G. J. Foran for assistance with XANES experiments at the Australian National Beamline Facility that provided the foundations for the present study. We also thank the Australian Research Council (DP0450252), the Access to Major Research Facilities Programme (funded by the Commonwealth of Australia) and the Natural Environment Research Council for financial support. GeoSoilEnviroCARS is supported by the US National Science Foundation (EAR-0622171) and the US Department of Energy (DE-FG02-94ER14466). Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract no. W-31-109-Eng-38.

Author Contributions A.J.B. and H.StC.O'N. prepared the synthetic samples and collected the XANES spectra, with assistance from M.N. and S.R.S. The natural samples were prepared and characterised by L.V.D., who identified the importance of determining $Fe^{3+}/\Sigma Fe$ in komatiite melt inclusions. A.J.B. interpreted the spectra and produced the manuscript with significant contributions from H.StC.O'N. and L.V.D.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.J.B. (a.berry@imperial.ac.uk).

METHODS

For the acquisition of XANES spectra, samples were mounted on Kapton tape and viewed from behind (downstream) using a long-working-distance objective located on the beam axis, which was removed during spectral acquisition. The optics allowed the beam to be accurately positioned on selected parts of a sample. To be certain that fluorescence from the host olivine did not contribute to the iron spectrum, the Kα intensity of nickel was monitored; nickel is an order of magnitude more abundant in the olivine than the inclusion. The excitation energy was selected using a Si(111) double-crystal monochromator and focused with Kirkpatrick–Baez mirrors. The spectral resolution at the iron K edge is ~1.6 eV. Spectra were recorded from 7,050 to 7,300 eV using step sizes of 2 eV from 7,050 to 7,100 eV, 0.2 eV from 7,100 to 7,160 eV, and 3 eV above 7,160 eV, in both transmission (ion chamber detector) and fluorescence (16-element germanium array detector) modes, although only fluorescence spectra are reported here. Spectra recorded sequentially as a function of time were identical and there was no evidence of any beam-induced changes in oxidation state. An iron foil spectrum was simultaneously acquired using the transmitted beam, allowing the energy of every sample spectrum to be individually calibrated by defining the first-derivative peak of the iron foil spectrum to be 7,112.0 eV. The sloping baseline between 7,050 and 7,100 eV was fitted to a linear function and the spectra 'straightened' before normalization to the average fluorescence above 7,200 eV.

The MORB sample (WW10-113-16) was dredged from the southeast Indian ridge at 105.22° E, 48.75° S during the WW10 voyage of the RV *Melville* in 1995.

LETTERS

Pairs of cooperating cleaner fish provide better service quality than singletons

Redouan Bshary[1], Alexandra S. Grutter[2], Astrid S. T. Willener[1] & Olof Leimar[3]

Service providers may vary service quality depending on whether they work alone or provide the service simultaneously with a partner. The latter case resembles a prisoner's dilemma[1–4], in which one provider may try to reap the benefits of the interaction without providing the service. Here we present a game-theory model based on the marginal value theorem[5], which predicts that as long as the client determines the duration, and the providers cooperate towards mutual gain, service quality will increase in the pair situation. This prediction is consistent with field observations and with an experiment on cleaning mutualism, in which stable male–female pairs of the cleaner wrasse *Labroides dimidiatus* repeatedly inspect client fish jointly. Cleaners cooperate by eating ectoparasites[6] off clients but actually prefer to cheat and eat client mucus[7]. Because clients often leave in response to such cheating, the benefits of cheating can be gained by only one cleaner during a pair inspection. In both data sets, the increased service quality during pair inspection was mainly due to the smaller females behaving significantly more cooperatively than their larger male partners. In contrast, during solitary inspections, cleaning behaviour was very similar between the sexes. Our study highlights the importance of incorporating interactions between service providers to make more quantitative predictions about cooperation between species.

Many cooperative interactions can be seen as an exchange of goods, services or commodities between two classes of traders[8–10]. Here we investigated traders that provide a service to a second class of traders, such as an ant partner species—for example, lycaenid butterfly larvae—providing a sugary solution to ants[11], rhizobial bacteria fixing nitrogen for leguminous plants[12] or cleaner fish removing ectoparasites from client reef fish[13]. We have used the last example as our model system. Cleaners prefer the mucus of some client species more than gnathiid isopods[7], the most commonly found ectoparasites of reef fishes[14]. Clients use various actions to make cleaners forage against their preference[15,16], the simplest form of control being to terminate the interaction by swimming off in response to a cheating bite[17]. Adult cleaners often live in pairs of a male and the largest female in his harem[18] and they commonly inspect larger clients simultaneously. Pair inspections result in cleaners facing a problem: a visiting client may leave after a cheat, even though only one cleaner was responsible for the cheating whereas the second cleaner cooperated. Hence, the cooperative cleaner loses a foraging opportunity owing to its partner's action, whereas the cheating cleaner gains a bite of mucus. We explored both mathematically and empirically how these pay-off asymmetries influence the service quality provided in paired compared with solitary inspections.

We explored a game in which one class of individuals provides a service (cleaners remove ectoparasites) to a second class of individuals (the clients). The service entails benefits to both the providers and the clients. However, the providers may also try to perform acts of exploitation (taking a bite of mucus) that yield a high immediate gain for them but are detrimental to the clients. Providers can decide on the level of service quality they provide, expressed as the frequency of exploits (the ratio of ectoparasites eaten per bite of mucus). Clients can decide on the duration of the interaction. In the cleaning mutualism, clients terminate interactions in response to cheating with about a 50% probability[19]. In the model, we assumed that clients have a 50% probability of terminating the interaction in response to an exploit. Under these conditions, we could find the optimal exploitation rate for the service provider. We assumed that the gains for providers from giving the service show diminishing returns with increasing duration (removal of ectoparasites leads to patch depletion). We used the expected value of all ectoparasites on a client as unit of benefit. We solve the problem for the cases in which: (1) one provider interacts with a client; (2) two providers cooperate with each other while interacting with the same client; and (3) two providers each try to maximise immediate individual gains while interacting with the same client.

For a singleton cleaner with a mucus bite rate (probability per unit time of taking a bite) of λ, the expected cleaning bout duration is $t = 2/\lambda$. The cleaner gain from ectoparasite removal is assumed to be $g_s(t)$, to which the benefit $2b$ of two bites per client on average is added. For large t, $g_s(t)$ approaches 1. For a cooperative pair of cleaners, each having the bite rate of $\lambda/2$, the expected cleaning duration is again t and the per capita gain is assumed to be $g_p(t)$, to which the per capita benefit b is added, because each cleaner has an equal chance of biting. We assume that $g_p(t) = g_s((1 + a)t)/2$, where a measures the added search efficiency when two cleaners service the client. Search efficiency doubles for $a = 1$, but interference between cleaners could lead to smaller values of a. As a fitness measure, we used the long-term gain rate. For an expected inter-client interval of t_0, the optimal bite rate for singletons and cooperative pairs (that is, the cleaning duration giving the maximum gain rate), satisfy the equations:

$$g'_s(t) = \frac{g_s(t) + 2b}{t + t_0} \quad (1)$$

$$g'_p(t) = \frac{g_p(t) + b}{t + t_0} \quad (2)$$

which are instances of the so-called marginal value theorem[5] (Fig. 1a). A cooperative pair maximises the total gain rate. For a non-cooperative pair, in which each cleaner attempts to maximise its own gain rate, an evolutionarily stable gain rate satisfies the equation

$$g'_s(t) = \frac{b}{t} + \frac{g_p(t) + b}{t + t_0} \quad (3)$$

provided that $t > 0$ holds. Immediate biting will be an evolutionarily

[1]University of Neuchâtel, Department of Zoology, Rue Emile-Argand 11 Case postale 158, 2009 Neuchâtel, Switzerland. [2]University of Queensland, School of Integrative Biology, St Lucia, Queensland 4072, Australia. [3]Stockholm University, Department of Zoology, SE-10691 Stockholm, Sweden.

stable strategy (ESS) for any $b > 0$, but there can also be an alternative ESS for which $t > 0$ (Fig. 1b; see Supplementary Information for derivation and analysis of equations (1), (2) and (3)). For small b there is an alternative non-cooperative ESS, for which the cleaning duration is smaller than the cooperative optimum (it approaches the cooperative optimum as b approaches zero). The client–cleaner mutualism involving pairs of cleaners is thus akin to a prisoner's dilemma, in which mutual cooperation yields higher pay-offs than mutual defection but defecting is dominant when cooperating in single rounds, at least for larger values of b. A general conclusion from our analyses is that cooperative pairs provide higher service quality than singletons (Fig. 1b; service quality is expressed as the proportion of removed ectoparasites per cheating bite) as long as there is added search efficiency for a pair ($a > 0$), whereas service quality would decrease (for $a < 1$) or stay the same (for $a = 1$) if each cleaner in the pair uses its singleton strategy (Fig. 1b).

We observed male–female pairs of cleaner fish in the field using standard interaction protocols (see Methods). We distinguished three situations—females inspecting alone, males inspecting alone, and pair inspection—and we compared client jolt rates (a response to cheating cleaners[19]) between the three situations. In the pair situation, we also distinguished between jolts caused by females and by males. In a laboratory experiment, we tested whether established cleaner fish pairs altered their willingness to feed against their preference (which translates into cooperative behaviour under natural conditions[7]) depending on whether they fed alone or with the partner. We considered three different possible outcomes: cleaners cheat more when cleaning in pairs (non-cooperative solution); cleaners behave in pairs how they behave when inspecting alone (solitary solution); or cleaners become more cooperative (cooperative solution). In the last case, client jolt rates during pair inspections should be lower than the sum of solitary male and female inspections, and the ratio of preferred to non-preferred items eaten in the experiment (prawn items eaten per fish flake item—called 'flake' hereafter[15]) should decrease in the pair situation. A tit-for-tat-like solution to the iterated prisoner's dilemma[1–4] would further suggest that males and females should equally adjust their behaviour during joint inspections.

Clients jolted significantly less frequently when interacting with a pair of cleaners compared to the sum of them interacting with females and males alone (Wilcoxon test, $n = 12$, $T = 13$, $P = 0.04$; Fig. 2). There were no significant differences in client jolt rates when interacting with either male or female alone (Wilcoxon test, $n = 12$, $T = 36$, $P > 0.05$), but females caused significantly less client jolts than males in the pair situation (Wilcoxon test, $n = 12$, $T = 6$, $P = 0.01$). Females also caused significantly less jolts when inspecting in a pair than when inspecting alone, whereas there was only a trend in the same direction for males (Wilcoxon test, females: $n = 12$, $T = 11$, $P = 0.023$; males: $n = 12$, $T = 16$, $P = 0.07$).

In the experiment, the ratio of (preferred) prawn items eaten per flake items eaten almost halved in pair trials compared with singleton trials (Fig. 3). There was no significant difference between male and female foraging behaviour when feeding on their own (Wilcoxon test, $n = 10$, 1 tie, resulting $n = 9$, $T = 19$, $P < 0.05$); both sexes ate significantly less prawn items per flake items in the pair situation compared with the singleton situation (Wilcoxon tests, males: $n = 10$, $T = 4$, $P = 0.017$; females: $n = 10$, $T = 0$, $P = 0.002$); and females ate significantly more against their preference than males in the pair situation (Wilcoxon test, $n = 10$, $T = 4$, $P = 0.017$). In the pair situation, males ate the prawn items significantly more often (on average 70%) than females (Wilcoxon rest, $n = 10$, 1 tie, resulting $n = 9$, $T = 2.5$, $P = 0.015$).

Males regularly chased females immediately after the removal of the plate on which food items were offered. Males were significantly more likely to chase females in trials in which the female had eaten the prawn item (median 100%) than in trials in which the male had eaten the prawn (median 50%; Wilcoxon test, $n = 10$, 2 ties, resulting $n = 8$, $T = 0$, $P = 0.008$).

Our model showed that two providers interacting simultaneously with the same client can face a situation similar to a prisoner's dilemma. The marginal value theorem allowed us to explore theoretically how diminishing returns influence the frequency of cheating. Exploring the time course of interactions may thus make models of cooperation more applicable to real life interactions, such as cooperative territorial defence in lionesses[20] or predator inspection in

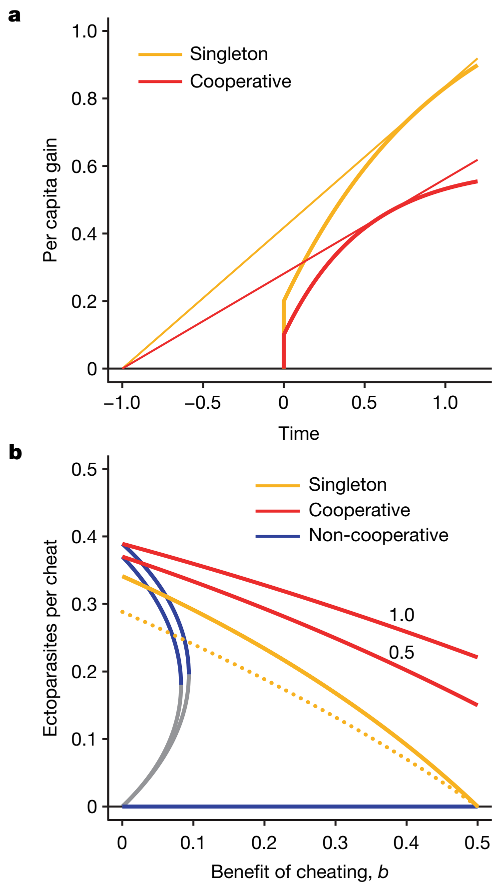


Figure 1 | The marginal value theorem for service providers. a, For singletons and cooperative pairs, the optimal cleaning duration occurs where a line from $(-t_0, 0)$ is tangent to the gain curve (inter-client interval $t_0 = 1$; $b = 0.1$). **b**, Service quality (measured as the proportion of removed ectoparasites per cheating bite) is higher for cooperative pairs than for singletons if pairs have higher search efficiency (cases with $a = 1$ and $a = 0.5$ are shown). Service quality is lower if each cleaner uses the singleton strategy (dotted curve, $a = 0.5$). Zero service quality (immediate biting) is a non-cooperative ESS for $b > 0$. For small b there is another ESS (the grey curve separates the basins of attraction of these ESSs).

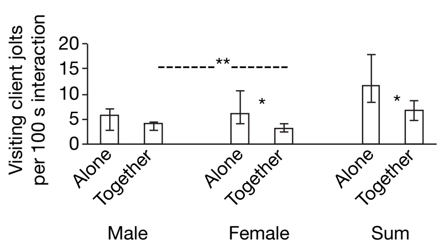


Figure 2 | Client jolt rates in response to singleton and pair inspections. Median and interquartiles of client jolt rates for 12 cleaner-fish pairs, with separate values for males when inspecting alone or with the female partner, for females when inspecting alone or with the male partner, for summed values of males and females inspecting alone, and total values for males and females inspecting together. Asterisk, $P < 0.05$; double asterisk, $P < 0.01$.

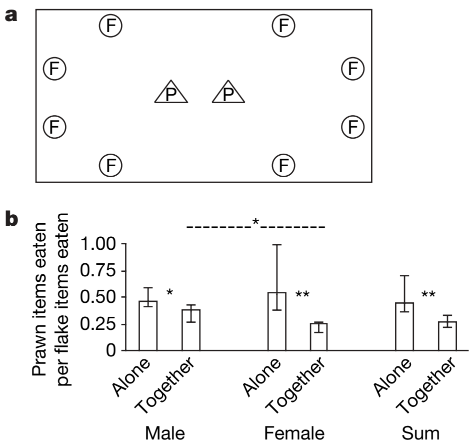


Figure 3 | Cleaner fish feed more against their preference when in pairs than when alone. a, The distribution of prawn items (P, in triangles) and of flake items (F, in circles) on the experimental Plexiglas plate is shown. **b**, Median and interquartile values of the ratio of prawn items eaten per flake items eaten for ten cleaner-fish pairs, with separate values for males when inspecting alone or with the female partner, for females when inspecting alone or with the male partner, for total values of males and females inspecting alone, and total values for males and females inspecting together. Asterisk, $P < 0.05$; double asterisk, $P < 0.01$.

fishes[21,22]. Our field observations and our aquarium experiment on cleaner fish pairs yielded consistent results that support the cooperative solution predicted by the model. Our results help to explain the observation that clients with choice options seem to preferentially visit stations with pairs of cleaner fish[23].

Although standard game theoretical analyses of iterated prisoner's dilemmas propose tit-for-tat-like solutions[1–4], these kinds of strategies are rarely found outside of humans[24]. Our two data sets superficially fit a tit-for-two-tats model in the sense that males cheated about twice as often as females during pair inspections. Alternatively, the unilateral aggression of the larger males towards females for cheating may function as punishment[25,26]. Although the importance of punishment as a control mechanism ensuring cooperation in symmetrical interactions has been disputed[27], we suggest that both empirical and theoretical future research should explore how dominance relationships affect solutions to the iterated prisoner's dilemma when punishment is a strategic option for the dominant individual.

METHODS SUMMARY

The model is described in Supplementary Information.

During field observations at Ras Mohammed National Park, Egypt, we noted cleaner identity (male, female or both), client species, duration of interaction and whether clients jolted in response to cleaner fish mouth contact (using methods as described previously[28]). Jolts provide a strong correlate of cleaners' cheating[19]. We calculated client jolt rates when interacting with cleaner pairs and when interacting with the male or the female only. In the pair situation, we also calculated the relative contribution of males and females to the total amount of client jolts.

In the laboratory experiment, cleaners were offered plates with two different food types, namely prawn and fish flakes mixed with prawn. The cleaners prefer prawn to flakes[16]. However, they could continue to eat as long as they ate only flakes, whereas eating a prawn item led to the immediate removal of the plate. Immediate reaction to prawn feeding was possible because the plate was attached to a lever held by the observer[16]. Cleaners thus had to eat against their preference to increase their foraging success. In the test trials, a brown Plexiglas plate (18 $\times$ 12 cm) with flake items placed within 8 black circles (each 1 cm diameter) drawn near the edges of the plate and prawn items placed within two black triangles in the centre (Fig. 3) was offered to cleaners when alone, and when paired in male–female pairs, and the sequence was balanced across individuals. For each cleaner and situation, the average ratio of prawn to flake items eaten in each round was calculated as a measure of how much cleaners were willing to feed against their preference. For the pair situation, we calculated the total number of prawn items eaten by males and by females, and we noted any aggression between partners.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 14 April; accepted 19 June 2008.

1. Axelrod, R. & Hamilton, W. D. On the evolution of co-operation. *Science* **211**, 1390–1396 (1981).
2. Nowak, M. A. & Sigmund, K. Tit for tat in heterogeneous populations. *Nature* **355**, 250–253 (1992).
3. Nowak, M. & Sigmund, K. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature* **364**, 56–58 (1993).
4. Dugatkin, L. A. *Cooperation Among Animals: An Evolutionary Perspective* (Oxford Univ. Press, 1997).
5. Charnov, E. L. Optimal foraging, the marginal value theorem. *Theor. Popul. Biol.* **9**, 129–136 (1976).
6. Grutter, A. S. Parasite removal rates by the cleaner wrasse *Labroides dimidiatus. Mar. Ecol. Prog. Ser.* **130**, 61–70 (1996).
7. Grutter, A. S. & Bshary, R. Cleaner wrasse prefer client mucus: support for partner control mechanisms in cleaning interactions. *Proc. R. Soc. Lond. B* **270** (Suppl.), S242–S244 (2003).
8. Noë, R., Van Schaik, C. P. & Van Hooff, J. A. R. A. M. The market effect: an explanation for pay-off asymmetries among collaborating animals. *Ethology* **87**, 97–118 (1991).
9. Noë, R. & Hammerstein, P. Biological markets. *Trends Ecol. Evol.* **10**, 336–339 (1995).
10. Bowles, S. & Hammerstein, P. in *Genetic and Cultural Evolution of Cooperation* (ed. Hammerstein, P.) 153–165 (MIT Press, 2003).
11. Pierce, N. E. *et al.* The ecology and evolution of ant associations in the Lycaenidae (Lepidoptera). *Annu. Rev. Entomol.* **47**, 733–771 (2002).
12. Kiers, E. T., Rousseau, R. A., West, S. A. & Denison, R. F. Host sanctions and the legume–rhizobium mutualism. *Nature* **425**, 78–81 (2003).
13. Côté, I. M. Evolution and ecology of cleaning symbioses in the sea. *Oceanogr. Mar. Biol. Ann. Rev.* **38**, 311–355 (2000).
14. Grutter, A. S. Relationship between cleaning rates and ectoparasite loads in coral reef fishes. *Mar. Ecol. Prog. Ser.* **118**, 51–58 (1995).
15. Bshary, R. & Grutter, A. S. Punishment and partner switching cause cooperative behaviour in a cleaning mutualism. *Biol. Lett.* **1**, 396–399 (2005).
16. Bshary, R. & Grutter, A. S. Image scoring and cooperation in a cleaner fish mutualism. *Nature* **441**, 975–978 (2006).
17. Johnstone, R. A. & Bshary, R. From parasitism to mutualism: partner control in asymmetric interactions. *Ecol. Lett.* **5**, 634–639 (2002).
18. Robertson, D. R. Social control of sex reversal in a coral-reef fish. *Science* **177**, 1007–1009 (1972).
19. Bshary, R. & Grutter, A. S. Asymmetric cheating opportunities and partner control in a cleaner fish mutualism. *Anim. Behav.* **63**, 547–555 (2002).
20. Heinsohn, R. & Packer, C. Complex cooperative strategies in group-territorial African lions. *Science* **269**, 1260–1262 (1995).
21. Milinski, M. TIT FOR TAT in sticklebacks and the evolution of cooperation. *Nature* **325**, 433–435 (1987).
22. Dugatkin, L. A. Do guppies play TIT FOR TAT during predator inspection visits? *Behav. Ecol. Sociobiol.* **23**, 395–399 (1988).
23. Bshary, R. & Schäffer, D. Choosy reef fish select cleaner fish that provide high-quality service. *Anim. Behav.* **63**, 557–564 (2002).
24. Hammerstein, P. *Why is Reciprocity so Rare in Social Animals? A Protestant Appeal* in *Genetic and Cultural Evolution of Cooperation* (ed. Hammerstein, P.) 83–93 (MIT Press, 2003).
25. Clutton-Brock, T. H. & Parker, G. A. Punishment in animal societies. *Nature* **373**, 209–215 (1995).
26. Clutton-Brock, T. H. & Parker, G. A. Sexual coercion in animal societies. *Anim. Behav.* **49**, 1345–1365 (1995).
27. Dreber, A., Rand, D. G., Fudenberg, D. & Nowak, M. A. Winners don't punish. *Nature* **452**, 348–351 (2008).
28. Bshary, R. & Würth, M. Cleaner fish *Labroides dimidiatus* manipulate client reef fish by providing tactile stimulation. *Proc. R. Soc. Lond. B* **268**, 1495–1501 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Egyptian Environmental Affairs Agency for permission to work at Ras Mohammed National Park, the Park Rangers and I. Riepl for local support, and M. Würth and A. Hubl for field assistance. We further thank the Lizard Island Research Station for their support concerning the experiments. R.B. was funded by the Natural Environment Research Council and by the Swiss Science Foundation. A.S.T.W. was funded by the Swiss Science Foundation. A.S.G. was funded by the Australian Research Council. O.L. was funded by the Swedish Research Council.

Author Contributions R.B. and A.S.G. were responsible for the experimental data; A.S.T.W. was responsible for the field data; O.L. was responsible for the model; and R.B., A.S.G. and O.L. contributed to the writing.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.B. (redouan.bshary@unine.ch).

METHODS

Field observations. Observations were made on 12 cleaner-fish pairs at Ras Mohammed National Park, Egypt from September 2006 to November 2006. Females were identified as the smaller individual in each pair. Each pair was observed for 400 min, with 200 min being focused on each individual. In client interactions with the pair that resulted in a jolt we noted which cleaner had caused it. In 32% of 1,296 cases, the identity could not be determined. These cases were attributed evenly to males and females for the statistical analyses. The statistical analyses required three steps of data processing. First, for each cleaning station, we identified the client species that had interacted with the pair and with the male and female only. We then determined for each species the jolt frequency when interacting with the male only or the female only, and when interacting with the cleaner pair the jolt frequency caused by the male and by the female. These values were used to calculate means of client jolt rates caused by males and by females when inspecting either alone or in pairs. In addition, we calculated the sums of client jolts when the male and female inspected alone and when the male and female inspected in pairs. The final analyses were conducted using Wilcoxon tests in which n was the number of cleaning stations.

Aquarium experiments. Experiments were conducted at the Lizard Island Research Station, Great Barrier Reef, Australia. Ten pairs of cleaners (total length of females 6.2–7.1 cm; males 7.0–8.9 cm; size difference within pairs 0.9–2.6 cm) were kept in aquaria of varying sizes (minimal size $50 \times 30 \times 25$ cm) for a minimum of 30 days before the experiments. All aquaria had running sea water and fish were provided with a PVC tube (1 cm diameter $\times$ 8 cm) for shelter. All cleaners were released after the experiment at the site of capture. Cleaners were trained to feed off Plexiglas plates. With the help of a temporary partition, they were individually familiarised with the main aspects of the experimental protocol—that is, that eating preferred prawn led to the removal of the plate. Within six rounds designed as learning trials, all cleaners ate flake items before eating a prawn item; the experiment was conducted afterwards.

In a sequence of 16 test trials distributed over 2 days, cleaners were confronted with the plate either alone or with their partner. The order of treatments was balanced: half of the individuals started with four pair trials, followed by eight singleton trials and another four pair trials; conversely, the other half started with four singleton trials, followed by eight pair trials and four singleton trials. During the day, a time interval of 40 min was maintained between trials. We also quantified how often male and female partners ate the prawn item in the pair situation and noted any aggressive behaviour of the partners (identity of aggressor and of victim) immediately after the removal of the plate.

LETTERS

Somatic and germline activating mutations of the ALK kinase receptor in neuroblastoma

Isabelle Janoueix-Lerosey[1,2], Delphine Lequin[1,2], Laurence Brugières[3], Agnès Ribeiro[4], Loïc de Pontual[5], Valérie Combaret[6], Virginie Raynal[1,2], Alain Puisieux[6,7,8], Gudrun Schleiermacher[1,2,9], Gaëlle Pierron[4], Dominique Valteau-Couanet[3], Thierry Frebourg[10], Jean Michon[9], Stanislas Lyonnet[5], Jeanne Amiel[5] & Olivier Delattre[1,2,4]

Neuroblastoma, a tumour derived from the peripheral sympathetic nervous system, is one of the most frequent solid tumours in childhood[1,2]. It usually occurs sporadically but familial cases are observed, with a subset of cases occurring in association with congenital malformations of the neural crest being linked to germline mutations of the *PHOX2B* gene[1–4]. Here we conducted genome-wide comparative genomic hybridization analysis on a large series of neuroblastomas. Copy number increase at the locus encoding the anaplastic lymphoma kinase (ALK)[5] tyrosine kinase receptor was observed recurrently. One particularly informative case presented a high-level gene amplification that was strictly limited to *ALK*, indicating that this gene may contribute on its own to neuroblastoma development. Through subsequent direct sequencing of cell lines and primary tumour DNAs we identified somatic mutations of the ALK kinase domain that mainly clustered in two hotspots. Germline mutations were observed in two neuroblastoma families, indicating that *ALK* is a neuroblastoma predisposition gene. Mutated ALK proteins were overexpressed, hyperphosphorylated and showed constitutive kinase activity. The knockdown of *ALK* expression in *ALK*-mutated cells, but also in cell lines overexpressing a wild-type *ALK*, led to a marked decrease of cell proliferation. Altogether, these data identify *ALK* as a critical player in neuroblastoma development that may hence represent a very attractive therapeutic target in this disease that is still frequently fatal with current treatments[6,7].

In a survey of the amplicons of neuroblastoma[8], we recently observed that high-level amplification (>10 copies) of *ALK*, a tyrosine kinase receptor (RTK) gene preferentially expressed in the central and peripheral nervous systems[5], was associated with *MYCN* amplification, the most frequent amplicon in neuroblastoma[1,2,9]. The expression level of *ALK* was strongly correlated with copy number[8], hence confirming and extending previous findings of *ALK* amplifications and overexpression in neuroblastoma[10–12].

To document more precisely *ALK* genomic status, we evaluated copy number variation at this locus in a series of 592 neuroblastomas that were investigated by bacterial artificial chromosome (BAC)-array comparative genomic hybridization (CGH). A total of 26 cases (4.4%) demonstrated higher than twofold copy number increases and an additional 135 cases (22.8%) presented lower level gains, therefore indicating that more that 25% of neuroblastomas harbour a significant copy number increase at the *ALK* locus (Fig. 1a). We also analysed the expression pattern of *ALK* across a series of paediatric tumours and normal tissues. This clearly showed that, whereas other tumour types occasionally expressed *ALK*, neuroblastoma consistently showed much higher levels of expression (Fig. 1b). In the course of this genomic analysis, a subset of cases without *MYCN* amplification was further studied by 100K single-nucleotide polymorphism (SNP) arrays. One particularly informative case was identified (NB-99) harbouring a unique and very high-level amplicon that was strictly limited to the *ALK* gene (Fig. 1c). The observation of such a focal and high-level amplicon suggested that *ALK* may have a role on its own in neuroblastoma.

Because RTK may be activated in human malignancies by genomic amplification, chromosome translocation or point mutations[13], we wondered whether *ALK* gene alterations, different from gene amplification, may be observed in neuroblastoma. We therefore searched for mutations in a first series of 28 cell lines. Single-nucleotide substitutions leading to changes of highly conserved amino acid residues were found and prompted us to extend the analysis to frozen tumour samples (115 cases). Altogether, the 16 identified mutations could be grouped into two main hotspots (Table 1). Notably, these changes were not observed in matched constitutional DNAs that were available for nine cases, hence demonstrating their somatic occurrence (Table 1 and Supplementary Fig. 1). In case CLB-GE that combined gene amplification and somatic mutation of the *ALK* gene, sequence analysis showed that the amplified and consequently overexpressed allele was the one with the F1174V change (Table 1), therefore suggesting that amplification and mutation may exert synergistic roles. In other cases, polymerase chain reaction with reverse transcription (RT–PCR) analysis indicated that both wild type and mutant alleles were expressed (Table 1). Although a F1174 mutation was observed in one primary tumour (NB-512), changes involving this amino acid were more frequent in cell lines, suggesting that they may provide a selective *in vitro* growth advantage (Table 1).

We also investigated the sequence of the kinase domain in six families presenting at least two siblings with neuroblastoma in the absence of malformation phenotype and in three cases associating neuroblastoma and Hirschprung's disease. A R1275Q heterozygous mutation was identified in one family. It was observed in the two affected children and inherited from the mother who was asymptomatic at 39 years of age (Fig. 1d). A R1192P heterozygous mutation was detected in another family. In this family, three out of six mutation carriers developed neuroblastic tumours, being ganglioneuroma in one case and stage 4s neuroblastomas in two other cases (Fig. 1d). No *ALK* mutation was detected in the other neuroblastoma families nor in Hirschprung's disease–neuroblastoma associations.

[1]Institut Curie, Centre de Recherche, [2]Inserm, U830, 26 rue d'Ulm, Paris F-75248, France. [3]Institut Gustave Roussy, Département de pédiatrie, 39 rue Camille Desmoulins, 94805 Villejuif, France. [4]Institut Curie, Unité de Génétique Somatique, Paris F-75248, France. [5]Département de Génétique, Université Paris Descartes, Faculté de Médecine et INSERM-U781, Hôpital Necker-Enfants Malades, 149, rue de Sèvres, 75743 Paris Cedex 15, France. [6]Centre Léon Bérard, FNCLCC, Laboratoire de Recherche Translationnelle. [7]Inserm, U590, [8]Université de Lyon, Lyon1, Institut des Sciences Pharmaceutiques et Biologiques, Lyon F-69008, France. [9]Institut Curie, Département de Pédiatrie, Paris F-75248, France. [10]Service de Génétique, CHU de Rouen et Inserm U614, Faculté de Médecine et de Pharmacie, 76183 Rouen Cedex, France.

We then investigated whether ALK mutation may result in a constitutive activation of the kinase activity. In agreement with expression data shown in Fig. 1b, the ALK protein was strongly expressed in neuroblastoma cell lines with particularly high levels in CLB-GE and CLB-GA that express the F1174V and R1275Q *ALK* mutant alleles, respectively (Fig. 2a). Surprisingly, in addition to the expected ALK band, the CLB-BAR cell line also expressed high levels of a lower molecular mass ALK protein (Fig. 2a). ALK proteins of CLB-GA, CLB-GE and KCNR, together with the truncated ALK of CLB-BAR, were highly reactive with the Y1586-specific ALK antibody that targets a phosphorylated residue associated with activation of the ALK protein[14]. In contrast, wild-type *ALK* of IMR32, CLB-BER, CLB-BAR and RH30 did not react with this antibody (Fig. 2a). ALK-specific immunoprecipitates from neuroblastoma cell lines were used for *in vitro* kinase assays. As compared to wild-type ALK, the mutant versions exhibited a much stronger *in vitro* kinase activity (Fig. 2b). In particular, the truncated ALK protein expressed by CLB-BAR harboured a very high kinase activity that was not detectable in the normal ALK isoform of this cell line. Altogether, these data show that ALK R1275Q and F1174L mutants and the truncated version of CLB-BAR harbour constitutive ALK kinase activity. Given the higher expression level of the mutant versions as compared to wild-type proteins, we cannot also exclude a role of the mutations in the stabilization of the proteins.

Knockdown of the *ALK* gene was then investigated using lentiviruses encoding *ALK*-specific or control small hairpin (sh) RNAs (Fig. 2c). Virally infected cells were plated at low density, maintained for puromycin selection for 10 days then stained with crystal violet. This showed that *ALK*-specific shRNA induced a specific and considerable inhibition of focus formation as compared to control shRNAs (Fig. 2d). Moreover, analysis of growth curves of *ALK*-knocked-down CLB-GA cells indicated a major reduction of cell proliferation as compared to control cells (Fig. 2e). The growth arrest was almost complete with the most efficient shRNAs (that is, sh5992 and sh5706). It was less pronounced with sh2854, in agreement with the only partial knockdown of *ALK* induced by this shRNA (Fig. 2c). Similar growth inhibition was observed for mutated CLB-MA or KCNR cells but also for IMR32 and 106C cells that harbour wild-type *ALK* sequences (Fig. 2e and data not shown).

Preliminary experiments addressed the issue of the downstream pathways that may be altered by ALK. Phosphorylated ERK1/2, MEK1/2, c-RAF and AKT were analysed on a panel of mutated and wild-type *ALK* cell lines (Supplementary Fig. 2). Apart from AKT, the phosphorylated form of which was detected in all cell lines, whatever the *ALK* status, phosphorylation of the other proteins was only occasionally observed. The knockdown of *ALK* led to a decreased expression of phospho-AKT in CLB-GA but not in IMR32 (Fig. 2f).

PF-2341066, a recently described inhibitor of c-MET activity that also shows ALK inhibitor activity[15], was tested on the growth of *ALK* mutated CLB-GA and CLB-MA cells, on *ALK* wild-type IMR32 cells, as well as on SK-N-AS cells that express neither *ALK* nor c-*MET*. As shown in Fig. 2g, the IC_{50} (half-maximal inhibitory concentration) values observed in the three former cell lines was significantly lower than the one for SK-N-AS.

Altogether, these data demonstrate mutations of the *ALK* gene in neuroblastoma that lead to constitutive kinase activity of the protein. A

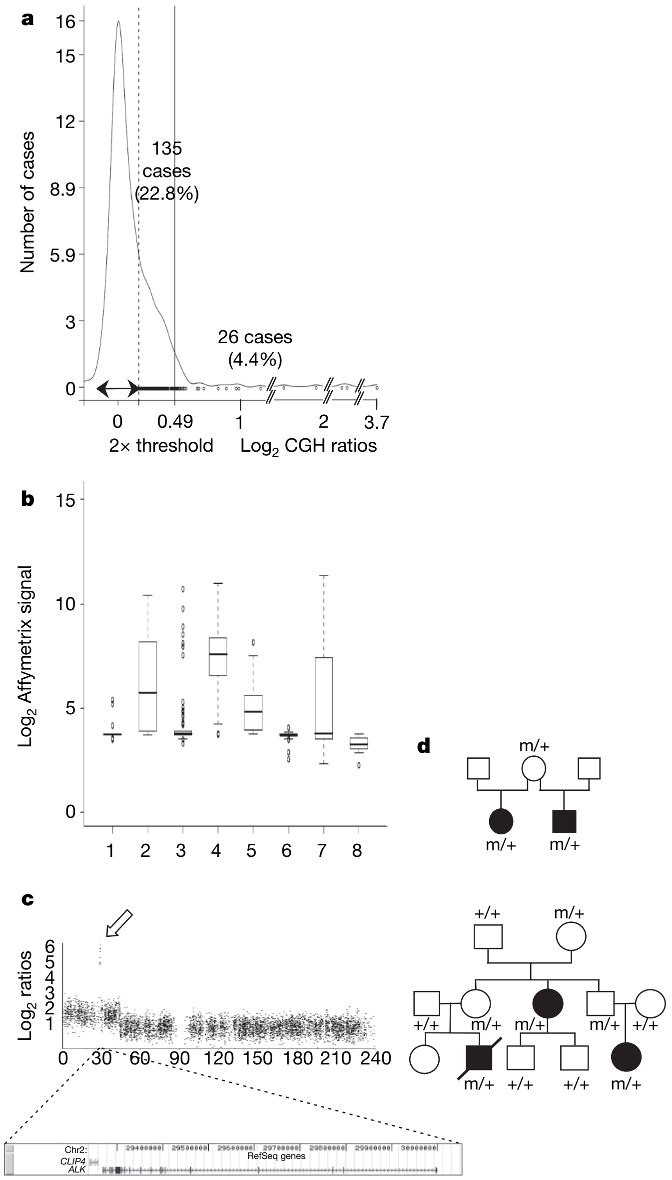


Figure 1 | Genetic characteristics of *ALK* in neuroblastoma. **a**, Distribution of the $\log_2$ values of tumour/control ratios at the RP11-328L16 (*ALK*) locus in 592 neuroblastomas analysed by BAC array-CGH. The double arrow shows the 99.5% confidence interval for normal copy number as defined by normal/normal array-CGH experiments. A total of 26 cases (4.4%) demonstrate $\log_2$ values higher than 0.49, a previously defined threshold for 2× copy number increase (vertical solid line)[27]. A total of 135 cases (22.8%) exhibit $\log_2$ ratios higher than the 99.5% confidence interval for normal ratios defined above (vertical dotted line) but lower than the 2× threshold. The values of RP11-328L16 across this series of samples are provided as Supplementary Data 1. **b**, Box-plots of expression levels of *ALK* (Affymetrix probe set 208212_s_at) across a series of 32 small round cell desmoplastic tumours (1), 39 Ewing samples (2), 75 medulloblastomas (3), 64 neuroblastomas (4), 26 malignant rhabdoid tumours (6), 122 rhabdomyosarcoma (7) and 11 teratomas (8). Normal tissue data (5) were downloaded from the GEO database (GSE3526). The box represents the middle 50% of data points. The upper and lower whiskers represent the highest and lowest observed values that are less than 1.5 box lengths from the ends of the box. The values of the 208212_s_at probe set across the complete series of samples is provided as Supplementary Data 2. **c**, SNP-array analysis of case NB-99. Only chromosome 2 is displayed. The *x* axis shows the physical distance (in megabases) from the telomere of the short arm. The *y* axis shows $\log_2$ ratios of copy number analysis. The arrow indicates the SNPs located within the *ALK* gene that show very high-level amplification (>25 copies). A UCSC scheme extending from SNPs flanking the amplicon is shown beneath the array-CGH data. No other amplicon, including *MYCN*, was observed and the *MYCN* coding sequence was wild type in this tumour. **d**, *ALK* mutations in two neuroblastoma families. In the top family, the R1275Q mutation (m) is detected in the daughter who developed a metastatic neuroblastoma at 15 years of age, in the son affected by a localized neuroblastoma at 3 years of age, and in the mother who is asymptomatic at 39. In the bottom family, the R1192P mutation is observed in the daughter who developed a ganglioneuroblastoma at 12 years of age and in the two grandchildren who developed stage 4s neuroblastomas at 3 and 4 months of age, respectively. The latter died of the disease.

Table 1 | *ALK* point mutations observed in neuroblastoma

Name	Sample	Stage	*MYCN* amplification	Genomic profile	*ALK* amplification*	Somatic analysis	Constitutional analysis	Type of mutation	Expressed allele(s)†
CLB-GE	Cell line	4	A	1p D, 17q G	A	Exon 23 F1174V (TTC>GTC)	WT	Somatic	Mutated
KCNR	Cell line	4	A	1p D, 17q G	NA	Exon 23 F1174L (TTC>TTA)	ND	Unknown	Both
CLB-GA	Cell line	4	NA	1p D, 11q D, 17q G	NA	Exon 25 R1275Q (CGA>CAA)	WT	Somatic	Both
SK-N-SH	Cell line	4	NA	17q G	NA	Exon 23 F1174L (TTC>TTA)	ND	Unknown	Both
N206	Cell line	4	A	1p D, 11q D, 17q G	NA	Exon 23 F1174L (TTC>TTA)	ND	Unknown	Both
CLB-BA	Cell line	4	A	1p D, 17q G	NA	Exon 23 F1174L (TTC>CTC)	WT	Somatic	Both
LAN1	Cell line	4	A	17q G	NA	Exon 23 F1174L (TTC>TTA)	ND	Unknown	ND
CLB-TR	Cell line	4	A	1p D, 17q G	ND	Exon 23 F1174L (TTC>CTC)	WT	Somatic	Both
CLB-MA	Cell line	4	A	1p D, 17q G	NA	Exon 23 F1174L (TTC>TTA)	WT	Somatic	Both
NB-212‡	Tumour	4s	NA	Numerical Alt	NA	Exon 25 R1275Q (CGA>CAA)	WT	Somatic	Both
NB-372	Tumour	1	NA	17q G	NA	Exon 25 R1275Q (CGA>CAA)	WT	Somatic	ND
NB-211	Tumour	2	NA	Numerical Alt	NA	Exon 25 R1275Q (CGA>CAA)	WT	Somatic	ND
NB-175	Tumour	2	NA	17q G	NA	Exon 25 Y1278S (TAC>TCC)	WT	Somatic	ND
NB-512	Tumour	4	NA	11q D, 17q G	NA	Exon 23 F1174C (TTC>TGC)	ND	Unknown	ND
NB-286	Tumour	4	NA	11 q D, 17q G	NA	Exon 25 R1275Q (CGA>CAA)	ND	Unknown	ND
NB-222	Tumour	2	NA	Numerical Alt	NA	Exon 25 R1275L (CGA>CTA)	ND	Unknown	Both

A, amplification; NA, no amplification; ND, not done; WT, wild type. The complete *ALK* coding sequence was investigated in all cell lines. Only the TK domain was screened in tumours. For the 'Genomic profile' column: D, deletion; G, gain; Numerical Alt, only numerical alterations.
* Gene amplification was investigated either by BAC-array CGH, Affymetrix 100K SNP arrays or semi-quantitative PCR.
† As deduced from sequence analysis of RT–PCR products.
‡ The *ALK* mutation was observed in the primary tumour and in a cutaneous nodule.

more exhaustive screen of the whole *ALK* coding sequence should now enable the determination of the exact frequency of *ALK* mutations in neuroblastoma. Recent data pointing out putative translocations involving *ALK* in neuroblastoma suggest that additional gene rearrangements of *ALK* remain to be identified[16]. With respect to clinical relationship, our first indications suggest that *ALK* mutations can be observed in tumours of various clinical stages and of various genomic profiles (Table 1). Moreover, a variety of neuroblastic tumours, from benign ganglioneuroma to very aggressive metastatic neuroblastoma, can be observed in the context of a germline *ALK* mutation. The observation that a number of mutation carriers remain asymptomatic indicates incomplete penetrance, a characteristic that is reminiscent of the *PHOX2B*-related neuroblastoma predisposition[4].

ALK is recurrently involved in gene fusions induced by chromosome translocations in different types of cancer including anaplastic large cell lymphomas[17], inflammatory myofibroblastic

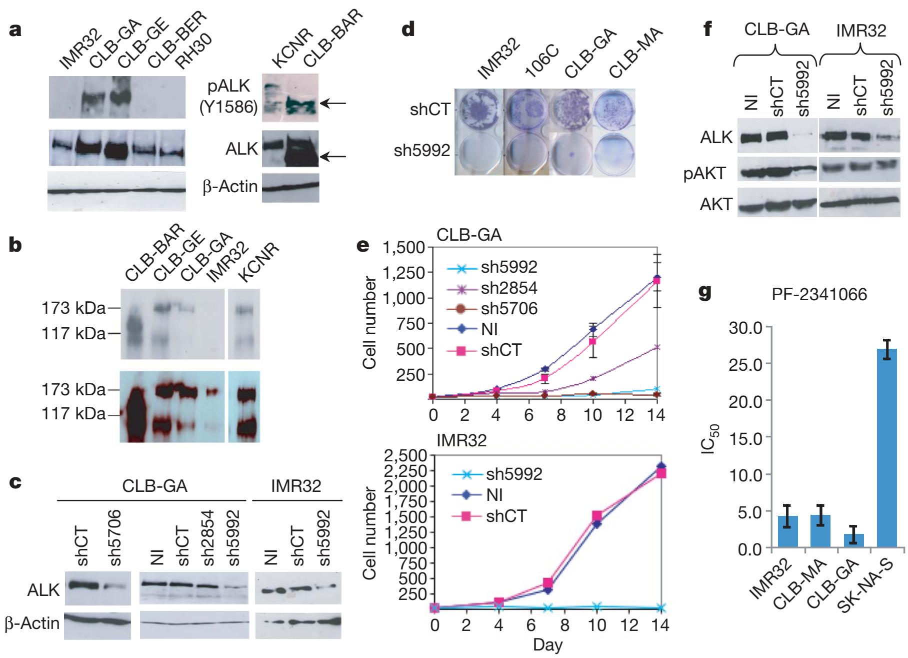


Figure 2 | Functional characterization of mutated ALK. **a**, Immunoblots of ALK and phosphorylated ALK receptor in various cell lines. RH30 is a rhabdomyosarcoma cell line. IMR32 and CLB-BER cells do not present *ALK* mutation. The arrow indicates the low molecular mass ALK protein that, in contrast to the wild-type ALK also expressed in CLB-BAR, strongly reacts with the Y1586 antibody. **b**, Strong kinase activity of ALK mutants. ^{32}P incorporated in ALK immunoprecipitates is visualized by autoradiography (top panel) before immunoblotting with an anti-ALK antibody (bottom panel). Mutated ALK from CLB-GE, CLB-GA and KCNR presents strong kinase activities whereas wild-type ALK from IMR32 barely exhibits any kinase activity. The abnormal ALK protein, but not the wild-type protein, of CLB-BAR demonstrates a strong kinase activity. **c**, Immunoblot after lentiviral-induced knockdown of *ALK* in the CLB-GA and IMR32 cell lines. NI, non-infected cells; shCT, cells infected with a control lentivirus. sh2854, sh5992 and sh5706 are *ALK*-specific small hairpin RNAs. **d**, Decreased focus formation of CLB-GA, IMR32, 106C and CLB-MA cells on *ALK* knockdown with sh5992 as compared to control short hairpin RNA (shCT). **e**, Decreased proliferation of CLB-GA and IMR32 cells on *ALK* knockdown. The number of living cells × 1,000 is reported along the *y* axis and the *x* axis indicates the days at which cells were counted. Means ± s.d. are indicated for CLB-GA. Duplicate experiments were performed for both cell lines. Similar results were obtained with the CLB-MA, KCNR and 106C cell lines (data not shown). **f**, Immunoblot of AKT and phosphorylated AKT in CLB-GA and IMR32 cells on *ALK* knockdown. **g**, PF-2341066 inhibition of cell growth in neuroblastoma cell lines expressing *ALK*. Mean IC_{50} values (μM) and s.d. of quadruplicate experiments are reported.

tumours[18] and a subset of lung cancers[19]. Most fusion proteins studied so far are characterized by constitutive activation of the ALK tyrosine kinase activity through spontaneous dimerization and induce cell transformation[5,20,21]. It remains to be documented that neuroblastoma-associated *ALK* mutations alone lead to cell transformation. However, mutations result in a constitutive activation of the kinase activity, a characteristic that is also observed after genetic changes of homologous domains, in particular of the activation loop, in other transforming RTK-encoding genes[22] (Supplementary Fig. 1). In many respects, *ALK* involvement in neuroblastoma is reminiscent of FLT3 contribution in acute myeloid leukaemia (AML). Indeed, whereas somatic mutations are only found in a subset of cases, most AMLs or neuroblastoma express high levels of *FLT3* or *ALK*, respectively[23]. In AML, there is some evidence that autocrine stimulation of the wild-type FLT3 receptor may confer a proliferative advantage to AML cells[23–25]. This cannot yet be tested in neuroblastoma as the nature of the ALK ligand(s) is still a matter of debate[5], but this may constitute an attractive hypothesis to account for the proliferative role of wild-type *ALK*.

Our preliminary observations, together with recent reports showing the potential of pharmacological[5,6,15] or immunological[26] strategies to inhibit ALK, in particular in neuroblastoma cell lines[16], suggest that such approaches will probably represent key therapeutic opportunities for neuroblastoma that is still very frequently a deadly disease.

METHODS SUMMARY

Array-CGH experiments were conducted as previously described on the Institut Curie one megabase resolution BAC-array[8,27]. SNP analyses were based on the 100K Affymetrix SNP arrays.

For immunoblots, total ALK was detected using monoclonal ALK1 (anti-human CD246, DakoCytomation) or polyclonal (Zymed) antibodies. Y1586 phosphorylated ALK was revealed using the number 3343 antibody (Cell Signaling Technologies). For immunoprecipitation, 250 μg of sonicated lysates were incubated with polyclonal ALK antibodies (Zymed) for 2 h at 4 °C then with G Sepharose beads (Pharmacia) for 1 h at 4 °C. After washing, the immunoprecipitates bound to protein G Sepharose were resuspended in gel loading buffer and resolved by SDS–polyacrylamide gel electrophoresis (PAGE).

Control and *ALK*-specific small hairpin RNAs in the pLKO puromycin-resistant vector were purchased from Sigma (SIGMA Mission shRNA). HEK293 cells were used for packaging using the calcium phosphate method. For colony formation experiments, 3×10^4 transduced cells were plated at day 0 in 6-well dishes and stained with crystal violet at day 10. For proliferation analysis, 2×10^4 transduced cells were plated at day 0 in 24-well dishes. The number of living cells was counted at day 4, 7, 10 and 14 using a Vi-cell XR counter (Beckman Coulter).

For *in vitro* kinase assay, ALK immunoprecipitates were incubated with protein G Sepharose beads (Pharmacia), then beads were pelleted and washed in kinase buffer. Kinase reactions were performed for 15 min at 20 °C in 40 μl of kinase buffer supplemented with [γ-^{32}P]ATP (10 μCi, Perkin Elmer). Reactions were stopped by the addition of an equal volume of gel loading buffer. After SDS–PAGE, the gels were electrotransferred onto Hybond nitrocellulose membranes (Amersham Pharmacia Biotech), and the kinase reaction products were detected by autoradiography.

For PF-2341066 growth inhibition, a range of concentrations from 100 μM to 0.01 nM was tested. MTS assay was used to determine cell viability.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 May; accepted 28 August 2008.

1. Maris, J. M., Hogarty, M. D., Bagatell, R. & Cohn, S. L. Neuroblastoma. *Lancet* **369**, 2106–2120 (2007).
2. Brodeur, G. M. Neuroblastoma: biological insights into a clinical enigma. *Nature Rev. Cancer* **3**, 203–216 (2003).
3. Tonini, G. P., Longo, L., Coco, S. & Perri, P. Familial neuroblastoma: a complex heritable disease. *Cancer Lett.* **197**, 41–45 (2003).
4. Trochet, D. *et al.* Germline mutations of the paired-like homeobox 2B (PHOX2B) gene in neuroblastoma. *Am. J. Hum. Genet.* **74**, 761–764 (2004).
5. Chiarle, R., Voena, C., Ambrogio, C., Piva, R. & Inghirami, G. The anaplastic lymphoma kinase in the pathogenesis of cancer. *Nature Rev. Cancer* **8**, 11–23 (2008).
6. Galkin, A. V. *et al.* Identification of NVP-TAE684, a potent, selective, and efficacious inhibitor of NPM-ALK. *Proc. Natl Acad. Sci. USA* **104**, 270–275 (2007).
7. Li, R. & Morris, S. W. Development of anaplastic lymphoma (ALK) small-molecule inhibitors for cancer therapy. *Med. Res. Rev.* **28**, 372–412 (2008).
8. Fix, A. *et al.* Characterization of amplicons in neuroblastoma. High-resolution mapping using DNA microarrays, relationship with outcome, and identification of overexpressed genes. *Genes Chromosom. Cancer* **47**, 819–834 (2008).
9. Schwab, M. *et al.* Chromosome localization in normal human cells and neuroblastomas of a gene related to c-myc. *Nature* **308**, 288–291 (1984).
10. Lamant, L. *et al.* Expression of the ALK tyrosine kinase gene in neuroblastoma. *Am. J. Pathol.* **156**, 1711–1721 (2000).
11. Osajima-Hakomori, Y. *et al.* Biological role of anaplastic lymphoma kinase in neuroblastoma. *Am. J. Pathol.* **167**, 213–222 (2005).
12. Miyake, I. *et al.* Activation of anaplastic lymphoma kinase is responsible for hyperphosphorylation of ShcC in neuroblastoma cell lines. *Oncogene* **21**, 5823–5834 (2002).
13. Blume-Jensen, P. & Hunter, T. Oncogenic kinase signalling. *Nature* **411**, 355–365 (2001).
14. Rikova, K. *et al.* Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* **131**, 1190–1203 (2007).
15. Zou, H. Y. *et al.* An orally available small-molecule inhibitor of c-Met, PF-2341066, exhibits cytoreductive antitumor efficacy through antiproliferative and antiangiogenic mechanisms. *Cancer Res.* **67**, 4408–4417 (2007).
16. McDermott, U. *et al.* Genomic alterations of anaplastic lymphoma kinase may sensitize tumors to anaplastic lymphoma kinase inhibitors. *Cancer Res.* **68**, 3389–3395 (2008).
17. Morris, S. W. *et al.* Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science* **263**, 1281–1284 (1994).
18. Griffin, C. A. *et al.* Recurrent involvement of 2p23 in inflammatory myofibroblastic tumors. *Cancer Res.* **59**, 2276–2280 (1999).
19. Soda, M. *et al.* Identification of the transforming *EML4-ALK* fusion gene in non-small-cell lung cancer. *Nature* **448**, 561–566 (2007).
20. Chiarle, R. *et al.* NPM-ALK transgenic mice spontaneously develop T-cell lymphomas and plasma cell tumors. *Blood* **101**, 1919–1927 (2003).
21. Fujimoto, J. *et al.* Characterization of the transforming activity of p80, a hyperphosphorylated protein in a Ki-1 lymphoma cell line with chromosomal translocation t(2;5). *Proc. Natl Acad. Sci. USA* **93**, 4181–4186 (1996).
22. Hubbard, S. R. Juxtamembrane autoinhibition in receptor tyrosine kinase. *Nature Rev. Mol. Cell Biol.* **5**, 464–471 (2004).
23. Renneville, A. *et al.* Cooperating gene mutations in acute myeloid leukaemia: a review of the literature. *Leukemia* **22**, 915–931 (2008).
24. Gilliland, G. & Griffin, G. The roles of FLT3 in hematopoiesis and leukaemia. *Blood* **100**, 1532–1542 (2002).
25. Fröhling, S. *et al.* Identification of driver and passenger mutations of FLT3 by high-throughput DNA sequence analysis and functional assessment of candidate alleles. *Cancer Cell* **12**, 501–513 (2007).
26. Chiarle, R. *et al.* The anaplastic lymphoma kinase is an effective oncoantigen for lymphoma vaccination. *Nature. Med.* **14**, 676–680 (2008).
27. Vincent-Salomon, A., Raynal, V., Lucchesi, C., Gruel, N. & Delattre, O. ESR1 gene amplification in breast cancer: a common phenomenon? *Nature Genet.* **40**, 809 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to F. Moreau-Gachelin, I. Gallais, D. Surdez, F. Tirode, A. Fix, F. Bourdeaut, A. Almeida, C. Lucchesi, S. Roman Roman, B. Bressac, J. Bénard and G. Vassal for their critical help. We thank C. Decraene, D. Gentien and B. Albaud from the translational department of Institut Curie for profiling the paediatric tumours, and P. Rosa and E. Barillot for the development of bioinformatic tools. We thank M. Lathrop and the Centre National de Génotypage for the Affymetrix 100K SNP analysis and A. Chompret for collecting families. This work was supported by grants from the Agence Nationale pour la Recherche, the Institut National du Cancer, the Ligue Nationale contre le Cancer (Equipe labellisée and CIT project), the APAESIC (Association des Parents et des Amis des Enfants Soignés à l'Institut Curie), the Association Hubert Gouin, les amis de Claire, Les Bagouz à Manon and Enfance et Santé. A.P. and V.C. are supported by the Comité de l'Ain of the Ligue Nationale contre le Cancer.

Author Contributions I.J.-L., D.L., A.R., L.P., V.C. and V.R. generated the data; I.J.-L., G.P., A.P., J.M., J.A., S.L. and O.D. made the study design and follow-up; L.B., V.C., A.P., G.S., D.V.-C., T.F., S.L. and J.A. contributed biological materials that were used in this study.

Author Information Microarray data have been submitted to the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) public database. The accession numbers for gene expression profiles of neuroblastoma samples and SNP data for case NB-99 displayed in Fig. 1c are GSE12460 and GSE12461, respectively. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to O.D. (olivier.delattre@curie.fr).

METHODS

Array-CGH experiments. Experiments were conducted as previously described on the Institut Curie one megabase resolution BAC-array[8,27]. The RP11-328L16 BAC contains most of the *ALK* gene. SNP analyses were based on the 100K Affymetrix SNP arrays (GeneChip 50K array Hind and Xba).

DNA sequencing was performed using the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit and sequences were analysed using SeqScape software (Applied Biosystems).

Immunoblots. An antibody (number 3343, Cell Signaling Technologies) directed against the phosphorylated Y1586 residue was used to reveal phosphorylated ALK. Total ALK was detected using the monoclonal ALK1 (anti-human CD246, DakoCytomation) or polyclonal (Zymed) antibodies.

RNA interference. Control and *ALK* specific small hairpin RNAs sh2854 (5′-CCGGACCCAAATCAAGAAACCTGTTCTCGAGAACAGGTTTCTTGATTTGGGTTTTTT-3′), sh5992 (5′-CCGGGTGATAAATACAAGGCCCAGACTCGAGTCTGGGCCTTGTATTTATCACTTTTT-3′) and sh5706 (5′-CCGGGAGCTGGTCATTACGAGGATACTCGAGTATCCTCGTAATGACCAGCTCTTTTT-3′) in the pLKO puromycin-resistant vector were purchased from Sigma (SIGMA Mission shRNA). HEK293 cells were used for packaging using the calcium phosphate method. Viral supernatants were supplemented with 8 μg ml^{-1} polybrene and incubated with neuroblastoma cells for 7 h. Selection was then applied with puromycin at 600 ng ml^{-1} for CLB-GA, 800 ng ml^{-1} for IMR32 and 1 μg ml^{-1} for 106C and CLB-MA. For colony formation experiments, 3×10^4 transduced cells were plated at day 0 in 6-well dishes and stained with crystal violet at day 10. For proliferation analysis, 2×10^4 transduced cells were plated at day 0 in 24-well dishes. The number of living cells was counted at day 4, 7, 10 and 14 using a Vi-cell XR counter (Beckman Coulter).

Immunoprecipitation. Cells were lysed in ice-cold lysis buffer (10 mM Tris-Hcl, pH 7.4, 150 mM NaCl, 5 mM EDTA, 20 mM NaF, 25 mM β-glycerophosphate, 1 mM Na-pyrophosphate, 10% glycerol, 1% NP40, 0.25% Na deoxycholate, 1 mM Na_3VO_4, 1× protease inhibitor; Roche). Lysates were sonicated then centrifuged for 30 min at 13,000 r.p.m. at 4 °C. Supernatants were recovered and protein concentration was determined by the Bradford assay. A total of 250 μg were incubated with polyclonal ALK antibodies (Zymed) for 2 h at 4 °C then with G Sepharose beads (Pharmacia) for 1 h at 4 °C. The immunoprecipitates bound to protein G Sepharose were washed twice with lysis buffer (without Na-deoxycholate) then re-suspended in gel loading buffer and resolved by SDS–PAGE.

***In vitro* kinase assay.** After incubation of ALK immunoprecipitates with protein G Sepharose beads (Pharmacia), the beads were pelleted and washed twice with buffer containing 10 mM Tris-Hcl, pH 7.4, 150 mM NaCl, 5 mM EDTA, 20 mM NaF, 25 mM β-glycerophosphate, 1 mM Na-pyrophosphate and 10% glycerol then washed with kinase buffer (20 mM HEPES, pH 7.4, 10 mM $MnCl_2$ and 0.5 mM Na_3VO_4). Kinase reactions were performed for 15 min at 20 °C in 40 μl of kinase buffer supplemented with [γ-^{32}P]ATP (10 μCi, Perkin Elmer). Reactions were stopped by the addition of an equal volume of gel loading buffer. After SDS–PAGE, the gels were electrotransferred onto Hybond nitrocellulose membranes (Amersham Pharmacia Biotech), and the kinase reaction products were detected by autoradiography.

PF-2341066 growth inhibition. A range of concentration from 100 μM to 0.01 nM was tested. MTS assay was used to determine cell viability.

LETTERS

Oncogenic mutations of ALK kinase in neuroblastoma

Yuyan Chen[1,2,3]*, Junko Takita[1,2,3]*, Young Lim Choi[4]*, Motohiro Kato[1,3], Miki Ohira[5], Masashi Sanada[2,3,6], Lili Wang[2,3,6], Manabu Soda[4], Akira Kikuchi[7], Takashi Igarashi[1], Akira Nakagawara[5], Yasuhide Hayashi[8], Hiroyuki Mano[4,6] & Seishi Ogawa[2,3,6]

Neuroblastoma in advanced stages is one of the most intractable paediatric cancers, even with recent therapeutic advances[1]. Neuroblastoma harbours a variety of genetic changes, including a high frequency of *MYCN* amplification, loss of heterozygosity at 1p36 and 11q, and gain of genetic material from 17q, all of which have been implicated in the pathogenesis of neuroblastoma[2–5]. However, the scarcity of reliable molecular targets has hampered the development of effective therapeutic agents targeting neuroblastoma. Here we show that the anaplastic lymphoma kinase (ALK), originally identified as a fusion kinase in a subtype of non-Hodgkin's lymphoma (NPM–ALK)[6–8] and more recently in adenocarcinoma of lung (EML4–ALK)[9,10], is also a frequent target of genetic alteration in advanced neuroblastoma. According to our genome-wide scans of genetic lesions in 215 primary neuroblastoma samples using high-density single-nucleotide polymorphism genotyping microarrays[11–14], the *ALK* locus, centromeric to the *MYCN* locus, was identified as a recurrent target of copy number gain and gene amplification. Furthermore, DNA sequencing of *ALK* revealed eight novel missense mutations in 13 out of 215 (6.1%) fresh tumours and 8 out of 24 (33%) neuroblastoma-derived cell lines. All but one mutation in the primary samples (12 out of 13) were found in stages 3–4 of the disease and were harboured in the kinase domain. The mutated kinases were autophosphorylated and displayed increased kinase activity compared with the wild-type kinase. They were able to transform NIH3T3 fibroblasts as shown by their colony formation ability in soft agar and their capacity to form tumours in nude mice. Furthermore, we demonstrate that downregulation of *ALK* through RNA interference suppresses proliferation of neuroblastoma cells harbouring mutated *ALK*. We anticipate that our findings will provide new insights into the pathogenesis of advanced neuroblastoma and that *ALK*-specific kinase inhibitors might improve its clinical outcome.

To identify oncogenic lesions in neuroblastoma, we performed a genome-wide analysis of primary tumour samples obtained from 215 neuroblastoma patients using high-density single-nucleotide polymorphism (SNP) arrays (Affymetrix GeneChip 250K *Nsp*I) (Supplementary Table 1). Twenty-four neuroblastoma-derived cell lines were also analysed (Supplementary Table 2). Interrogating over 250,000 SNP sites, this platform permits the identification of copy number changes at an average resolution of less than 12 kilobases (kb)[13,14].

Analysis of this large number of samples, consisting of varying disease stages, permitted us to obtain a comprehensive registry of genomic lesions in neuroblastoma (Supplementary Figs 1 and 2). A gain of chromosomes, often triploid or hyperploid (defined by mean copy number of >2.5), was a predominant feature of neuroblastoma genomes in the lower stages. Ploidy generally correlated with the clinical stage, where non-hyperploid cases were significantly associated with stage 4 disease ($P = 4.13 \times 10^{-5}$, trend test) (Supplementary Fig. 3 and Supplementary Table 3). 17q gains, frequently in multiple copies ($3 \leq$ copy number <5), were a hallmark of the neuroblastoma genome[4] and were found in most neuroblastoma cases. Copy number gains tended to spare chromosomes 3, 4, 10, 14 and 19 (Supplementary Figs 2 and 3). Notably, these chromosomes often had copy number losses including 1p (22.8%), 3p (8.8%), 4p (5.1%), 6q (7.0%), 10q (9.8%), 11q (19.5%), 14q (3.7%), 19p (7.4%) and 19q (5.1%), implicating the pathogenic role of 'relative' gene dosages.

After excluding known copy number variations, we identified a total of 28 loci undergoing high-grade amplifications (copy number ≥ 5) (Supplementary Table 4). These lesions fell into relatively small genomic segments, having a mean size of 361 kb, which accelerated the identification of gene targets in these regions (Supplementary Table 4 and Supplementary Fig. 4). The candidate gene targets included *TERT* (5p15.33), *HDAC3* (5q31.3), *IGF2* (11p15.1), *MYEOV* (11q13.3), *FGF7* (15q21.1) and *CDH13* (16q23.3). However, many of them were not recurrent but found only in a single case. Although the recurrent lesions were mostly explained by the amplification of *MYCN* at 2p24, as found in 50 out of 215 (23%) of the primary cases, we identified another peak of recurrent amplification at 2p23 (Fig. 1a), which consisted of amplicons in five primary cases and in one neuroblastoma-derived cell line, NB-1 (Supplementary Fig. 5). This peak was located at the centromeric margin of the common copy number gains in chromosome 2p, which was created by copy number gains in 109 samples mostly from non-hyperploid stage 4 cases. The minimum overlapping amplification was defined by the amplicons found in the NB-1 cell line (Supplementary Fig. 5) and contained a single gene, the anaplastic lymphoma kinase (*ALK*), which has previously been reported to be overexpressed in neuroblastoma cases[15]. Although five of the six samples showing *ALK* amplification also had *MYCN* amplification, one primary case (NT056) lacked a *MYCN* peak and the amplification was confined to the *ALK*-containing locus. In interphase fluorescent *in situ* hybridization (FISH) analysis of NB-1, *MYCN* and *ALK* loci were amplified in separate amplicons (Fig. 1b), indicating that the 2p23 amplicons containing *ALK* were unlikely to represent merely 'passenger' events of *MYCN* amplification but actively contributed to the pathogenesis of neuroblastoma.

Because an oncogene can be activated by gene amplification and/or mutation, to search for possible mutations we performed DNA heteroduplex formation analysis[16] and genomic DNA sequencing for the exons 20 to 28 of *ALK*, which encompass the juxtamembrane and kinase domains (Supplementary Table 5). In total, we identified eight nucleotide changes in 21 neuroblastoma samples, 13 out of 215

[1]Department of Pediatrics, [2]Cell Therapy and Transplantation Medicine, [3]Cancer Genomics Project, Graduate School of Medicine, The University of Tokyo, Tokyo 113-8655, Japan. [4]Division of Functional Genomics, Jichi Medical University, Toguchi 329-0498, Japan. [5]Division of Biochemistry, Chiba Cancer Center Research Institute, Chiba 260-8717, Japan. [6]Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Saitama, 332-0012, Japan. [7]Division of Hematology/Oncology, Saitama Children's Medical Center, Saitama 339-8551, Japan. [8]Gunma Children's Medical Center, Shibukawa 377-8577, Japan.
*These authors contributed equally to this work.

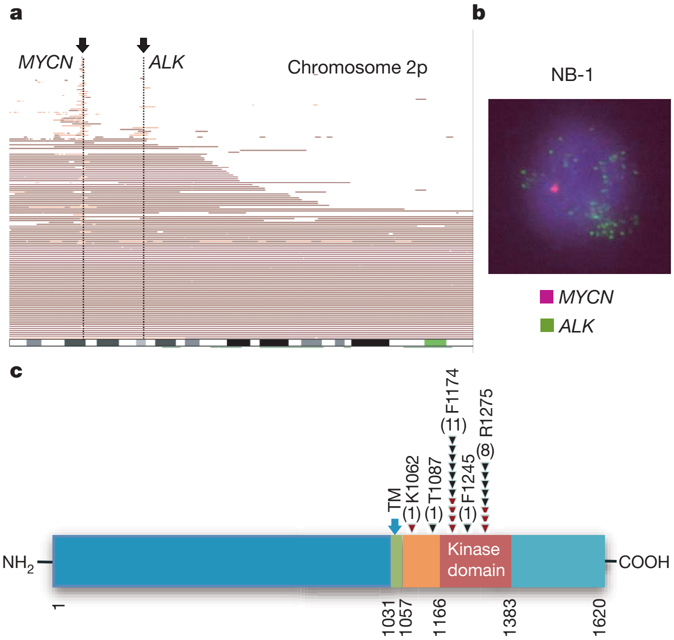


Figure 1 | Common 2p gains/amplifications and *ALK* mutations in neuroblastoma samples. **a**, Recurrent copy number gains on the 2p arm. High-grade amplifications are shown by light-red horizontal lines, whereas simple gains are shown by dark-red lines. Two common peaks of copy number gains and amplifications in the *MYCN* and *ALK* loci are indicated by arrows. The cytobands in 2p are shown at the bottom. **b**, Interphase FISH analysis of NB-1 showing high-grade amplification of *MYCN* (red) and *ALK* loci (green). The amplified *MYCN* locus appears as a single large signal. **c**, Distribution of the eight *ALK* mutations found in 21 neuroblastoma samples. The positions of the mutated amino acids are indicated by black (primary samples) and red (cell lines) arrowheads. The number of mutations at each site is shown at the top of the arrowheads. TM, transmembrane.

(6.1%) primary samples and 8 out of 24 (33%) cell lines, which resulted in seven types of amino acid substitutions at five different positions (Table 1 and Supplementary Fig. 6). They were not found in either the genomic DNA collected from 50 healthy volunteers or in the SNP databases at the time of preparing this manuscript. In fact, somatic origins of missense changes were confirmed in 9 out of 13 primary cases, for which DNA was obtained from the peripheral blood or the tumour-free bone marrow specimens (Supplementary Fig. 6). On the other hand, T1087I (ACC>ATC), found in case NT126, had a germline origin and thus it could not be determined whether the T1087I change was a rare non-functional polymorphism or represented a pathogenic germline mutation. For other changes found in three primary cases (NT128, NT217 and NT218) and cell lines, normal DNA was not available but they were likely to represent oncogenic mutations because they were identical to common somatic changes (F1174L or R1275Q) or shown to have oncogenic potential in functional assays (K1062M).

Most mutations occurred within the kinase domain (20 out of 22 or 91%), which clearly showed two mutation hotspots at F1174 and R1275 (Fig. 1c). A neuroblastoma-derived cell line, SJNB-2, had a homozygous *ALK* mutation of R1275Q, which was probably due to uniparental disomy of chromosome 2 (Supplementary Fig. 7a). Another case (NT074) harboured two different mutations, F1174L and R1275Q, but it remains to be determined whether both are on the same allele. ALK mutations within the kinase domain occurred at amino acid positions that are highly conserved across species and during molecular evolution (Supplementary Figs 8 and 9). According to the conserved structure of other insulin receptor kinases we predicted that F1174 is located at the end of the Cα1 helix, whereas the other two are on the two β-sheets: before the catalytic loop (β6, F1245) and within the activation loop (β9, R1275) (Supplementary Fig. 7b, c)[17]. Thus, conformational changes due to amino acid substitutions at these positions might be responsible for the aberrant activity of the mutant kinases.

Table 1 | *ALK* mutations/amplifications in neuroblastoma samples

Sample	Age (months)	Stage	*MYCN**	Clinical outcome	Mutations/ amplifications	Nucleotide substitution	Origin of mutations
NT126	99	4	−	Dead	T1087I	ACC>ATC	Germ line
NT218	8	1	−	Alive	F1174L	TTC>TTG	ND
NT074	34	3	+	Dead	F1174L	TTC>TTA	Somatic
					R1275Q	CGA>CAA	
NT160	12	4	+	Dead	F1174L	TTC>TTA	Somatic
NT217	24	4	+	Dead	F1174L	TTC>TTA	ND
NT190	48	4	+	Alive	F1174L	TTC>TTA	Somatic
NT060	163	3	−	Alive	F1174C	TTC>TGC	Somatic
NT162	28	4	+	Dead	F1174V	TTC>GTC	Somatic
NT195	24	4	+	Alive	F1245L	TTC>TTG	Somatic
NT055	6	3	−	Alive	R1275Q	CGA>CAA	Somatic
NT128	8	4	−	Dead	R1275Q	CGA>CAA	ND
NT164	54	4	+	Dead	R1275Q	CGA>CAA	Somatic
NT200	133	4	−	Dead	R1275Q	CGA>CAA	Somatic
SCMC-N5†	-	-	+	-	K1062M	AAG>ATG	ND
SJNB-4†	-	-	+	-	F1174L	TTC>TTA	ND
LAN-1†	-	-	+	-	F1174L	TTC>TTA	ND
SCMC-N2†	-	-	+	-	F1174L	TTC>TTA	ND
SK-N-SH†	-	-	−	-	F1174L	TTC>TTA	ND
SJNB-2†‡	-	-	+	-	R1275Q	CGA>CAA	ND
LAN-5†	-	-	+	-	R1275Q	CGA>CAA	ND
TGW†	-	-	+	-	R1275Q	CGA>CAA	ND
NT204	12	1	+	Alive	Amplification	-	-
NT056	11	3	−	Dead	Amplification	-	-
NT071	36	3	+	Alive	Amplification	-	-
NT165	19	4	+	Dead	Amplification	-	-
NT169	7	4	+	Dead	Amplification	-	-
NB-1†	-	-	+	-	Amplification	-	-

ND, not determined.
* Presence (+) or absence (−) of *MYCN* amplification in FISH analysis. All cases where there was an absence of *MYCN* amplification (−) were also checked for possible *MYCN* mutations by sequencing of all *MYCN* exons, but no *MYCN* mutations were identified.
† Cell lines.
‡ Homozygous mutation.

ALK mutation highly correlated with *MYCN* amplification ($P = 1.55 \times 10^{-4}$, Fisher's exact test; Supplementary Table 6) where 14 out of 21 mutations coexisted with *MYCN* amplification. Regardless of the status of *MYCN* amplification, 12 of the 13 mutations were found in patients with advanced stage neuroblastoma (Table 1). However, whereas *MYCN* amplification and stage 4 were significant risk factors for poor survival, the mutation/amplification status of *ALK* was not likely to have a major impact on survival (Supplementary Fig. 10 and Supplementary Table 7), although the statistical power of the current analysis was largely limited in order to detect a marginal hazard.

To evaluate the impact of ALK mutations on kinase activity, we generated Flag-tagged constructs of ALK and its mutants, F1174L and K1062M, which were stably expressed in NIH3T3 cells, and examined their phosphorylation status and *in vitro* kinase activity. The ALK mutants stably expressed in NIH3T3 cells were phosphorylated according to western blot analysis using an antibody specific for phosphorylated ALK (anti-pY1604) and a PY20 blot after anti-Flag immunoprecipitation of the mutant kinases (Fig. 2a), whereas the wild-type kinase was not phosphorylated. The immunoprecipitated ALK mutants also showed increased tyrosine kinase activity *in vitro* when compared with wild-type ALK. This was shown using both a universal substrate for tyrosine kinase (poly-GluTyr) and the synthetic YFF peptide[18], which was derived from a sequence of the activation loop of *ALK* (Fig. 2b, c). In accordance with these findings, downstream molecules of ALK signalling including AKT, STAT3 and ERK[15] were activated in cells expressing mutant ALK, as shown by their increased phosphorylation (Fig. 2d).

Next, we investigated the oncogenic potential of these mutants. NIH3T3 cells stably expressing mutant kinases showed increased colony formation in soft agar compared with the wild-type protein (Fig. 3a and Supplementary Fig. 11). The tumorigenicity of these ALK mutants was further assayed by injecting 1.0×10^7 NIH3T3 cells into nude mice. The NIH3T3 cells transfected with the ALK mutants showed focus-forming capacity and developed subcutaneous tumours (6 out of 6 inoculations) 21 days after inoculation, whereas the mock and wild-type ALK-transfected cells did not (0 out of 6 inoculations) (Fig. 3b, c). Finally, we examined the effect of ALK inhibition on the proliferation of neuroblastoma-derived cell lines. RNA interference (RNAi)-mediated *ALK* knockdown resulted in reduced cell proliferation of SK-N-SH cells harbouring the F1174L mutation, but the effects were less clear in wild-type ALK-expressing LAN-2 cells (Fig. 3d, e). Of particular interest is a recent report that 5 out of 17 neuroblastoma-derived cell lines, including SK-N-SH and NB-1, frequently showed high sensitivity to the specific ALK inhibitor TAE684 (ref. 19).

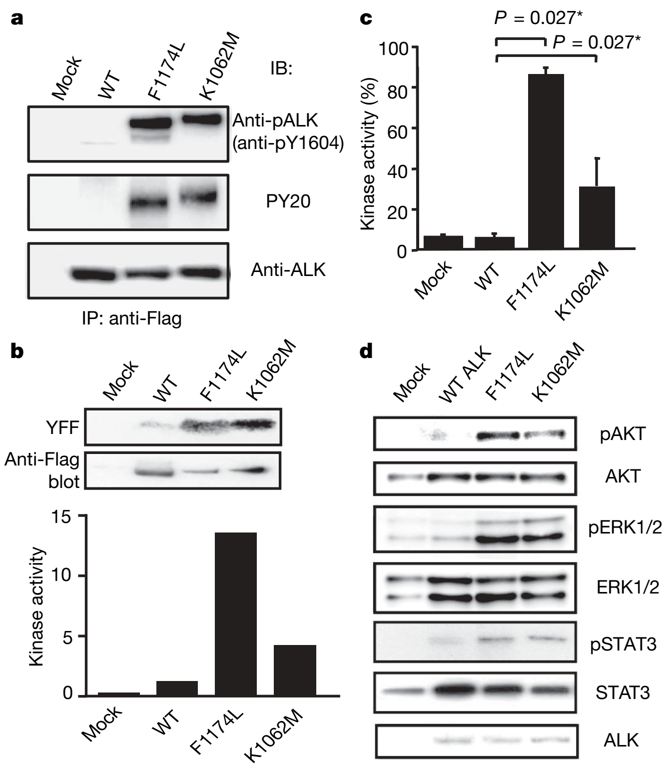


Figure 2 | Kinase activity of ALK mutants and their downstream signalling. **a**, Stably expressed ALK and its mutants (F1174L and K1062M) were immunoprecipitated with an anti-Flag antibody and subjected to western blot analysis with anti-pY1604 (upper panel) or PY20 (middle panel). An anti-ALK blot of precipitated kinases is also displayed (bottom panel). **b**, *In vitro* kinase assay for wild-type ALK kinase and its mutants using the synthetic YFF peptide as a substrate, where kinase activity is expressed as relative values to that for wild-type kinase based on the densities in the autoradiogram. **c**, Kinase activity was also assayed for the poly-GluTyr peptide. Significantly different measurements are indicated by asterisks with *P* values. Bars show mean (±s.d.) in three independent experiments. **d**, Western blot analyses of NIH3T3 cells expressing wild-type and mutant ALK for phosphorylated forms of AKT (pAKT), ERK (pERK1/2) and STAT3 (pSTAT3). The total amount of each molecule is also displayed (AKT, ERK1/2, and STAT3) together with an anti-ALK blot (ALK).

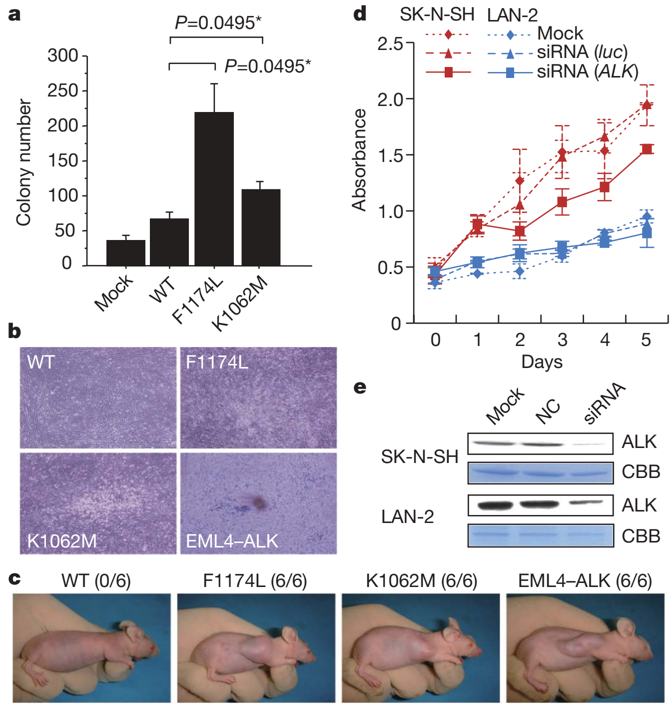


Figure 3 | Oncogenic role of ALK mutations. **a**, Colony assays for NIH3T3 cells stably expressing wild-type as well as mutant ALK (F1174L and K1062M). The average numbers of colonies in triplicate experiments are plotted and standard deviation is indicated. Results showing statistically significant differences as compared with experiments using wild-type ALK are indicated by asterisks with *P* values. **b**, **c**, NIH3T3 cells were transfected with wild-type and mutant ALK (F1174L, K1062M and EML4–ALK) and subjected to a focus forming assay (**b**) as well as an *in vivo* tumorigenicity assay in nude mice (**c**). **d**, Effect of RNAi-mediated ALK knockdown on cell proliferation in neuroblastoma cell lines expressing either the F1174L mutant (SK-N-SH) or wild-type ALK (LAN-2). Cell growth was measured using the Cell Counting Kit-8 after knockdown experiments using *ALK*-specific siRNAs (siRNA *ALK*), control siRNAs (siRNA *luc*), or mock experiments, where absorbance was measured in triplicate and averaged for each assay. To draw growth curves, the mean ± s.d. of the averaged absorbance in three independent knockdown experiments is plotted. **e**, Successful knockdown of ALK protein was confirmed by anti-ALK blots (ALK) using Coomassie brilliant blue G-250 (CBB) staining as loading controls. NC, control siRNA; siRNA, *ALK* siRNA.

Through the genome-wide analysis of genetic lesions in neuroblastoma, we identified novel oncogenic *ALK* mutations in advanced neuroblastoma. Combined with the cases having a high-grade amplification of the *ALK* gene, aberrant ALK signalling was likely to be involved in 11% (16 out of 151) of the advanced neuroblastoma cases. Because ALK kinase has been shown to be deregulated only in the form of a fusion kinase in human cancers, including lymphoma and lung cancer, the identification of oncogenic mutations in *ALK* not only increases our understanding of the molecular pathogenesis of advanced neuroblastoma, but also adds a new paradigm to the concept of 'ALK-positive human cancers' in that the mutated *ALK* kinases themselves might participate in human cancers. Our results again highlight the power of genome-wide studies to clarify the genetic lesions in human cancers[20–22]. Given that *ALK* mutations are preferentially involved in advanced neuroblastoma cases having a poor prognosis, our findings implicate that *ALK* inhibitors may improve the clinical outcome of children suffering from intractable neuroblastoma.

METHODS SUMMARY

Genomic DNA from 215 patients with primary neuroblastoma and 24 neuroblastoma-derived cell lines was analysed on GeneChip SNP genotyping microarrays (Affymetrix GeneChip 250K *Nsp*I). After appropriate normalization of mean array intensities, signal ratios were calculated between tumours and anonymous normal references in an allele-specific manner, and allele-specific copy numbers were inferred from the observed signal ratios based on the hidden Markov model using CNAG/AsCNAR software[13,14]. *ALK* mutations were examined by DNA heteroduplex analysis and/or genomic DNA sequencing[16]. Full-length cDNAs for mutant *ALK* were isolated by high-fidelity PCR and inserted into pcDNA3 and pMXS. The expression plasmids were transfected into NIH3T3 cells using Effectene Transfection Reagent (Qiagen) or by calcium phosphate methods[9]. Western blot analysis of mutant ALK kinases, *in vitro* kinase assays, and tumour formation assays in nude mice were performed as previously described[9]. This study was approved by the ethics boards of the University of Tokyo and of the Chiba Cancer Center Research Institute.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 June; accepted 28 August 2008.

1. Maris, J. M., Hogarty, M. D., Bagatell, R. & Cohn, S. L. Neuroblastoma. *Lancet* **369,** 2106–2120 (2007).
2. Maris, J. M. *et al.* Loss of heterozygosity at 1p36 independently predicts for disease progression but not decreased overall survival probability in neuroblastoma patients: a Children's Cancer Group study. *J. Clin. Oncol.* **18,** 1888–1899 (2000).
3. Attiyeh, E. F. *et al.* Chromosome 1p and 11q deletions and outcome in neuroblastoma. *N. Engl. J. Med.* **353,** 2243–2253 (2005).
4. Bown, N. *et al.* Gain of chromosome arm 17q and adverse outcome in patients with neuroblastoma. *N. Engl. J. Med.* **340,** 1954–1961 (1999).
5. Brodeur, G. M., Seeger, R. C., Schwab, M., Varmus, H. E. & Bishop, J. M. Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science* **224,** 1121–1124 (1984).
6. Shiota, M. *et al.* Anaplastic large cell lymphomas expressing the novel chimeric protein p80NPM/ALK: a distinct clinicopathologic entity. *Blood* **86,** 1954–1960 (1995).
7. Morris, S. W. *et al.* Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science* **263,** 1281–1284 (1994).
8. Fujimoto, J. *et al.* Characterization of the transforming activity of p80, a hyperphosphorylated protein in a Ki-1 lymphoma cell line with chromosomal translocation t(2;5). *Proc. Natl Acad. Sci. USA* **93,** 4181–4186 (1996).
9. Soda, M. *et al.* Identification of the transforming *EML4–ALK* fusion gene in non-small-cell lung cancer. *Nature* **448,** 561–566 (2007).
10. Rikova, K. *et al.* Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* **131,** 1190–1203 (2007).
11. Kennedy, G. C. *et al.* Large-scale genotyping of complex DNA. *Nature Biotechnol.* **21,** 1233–1237 (2003).
12. Matsuzaki, H. *et al.* Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nature Methods* **1,** 109–111 (2004).
13. Nannya, Y. *et al.* A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res.* **65,** 6071–6079 (2005).
14. Yamamoto, G. *et al.* Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of affymetrix single-nucleotide-polymorphism genotyping microarrays. *Am. J. Hum. Genet.* **81,** 114–126 (2007).
15. Osajima-Hakomori, Y. *et al.* Biological role of anaplastic lymphoma kinase in neuroblastoma. *Am. J. Pathol.* **167,** 213–222 (2005).
16. Donohoe, E. Denaturing high-performance liquid chromatography using the WAVE DNA fragment analysis system. *Methods Mol. Med.* **108,** 173–187 (2005).
17. Hu, J., Liu, J., Ghirlando, R., Saltiel, A. R. & Hubbard, S. R. Structural basis for recruitment of the adaptor protein APS to the activated insulin receptor. *Mol. Cell* **12,** 1379–1389 (2003).
18. Donella-Deana, A. *et al.* Unique substrate specificity of anaplastic lymphoma kinase (ALK): development of phosphoacceptor peptides for the assay of ALK activity. *Biochemistry* **44,** 8533–8542 (2005).
19. McDermott, U. *et al.* Genomic alterations of anaplastic lymphoma kinase may sensitize tumors to anaplastic lymphoma kinase inhibitors. *Cancer Res.* **68,** 3389–3395 (2008).
20. Garraway, L. A. *et al.* Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature* **436,** 117–122 (2005).
21. Mullighan, C. G. *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446,** 758–764 (2007).
22. Kawamata, N. *et al.* Molecular allelokaryotyping of pediatric acute lymphoblastic leukemias by high-resolution single nucleotide polymorphism oligonucleotide genomic microarray. *Blood* **111,** 776–784 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. P. Koeffler for critically reading and editing the manuscript. We also thank M. Matsumura, Y. Ogino, S. Ichimura, S. Sohma, E. Matsui, Y. Yin, N. Hoshino and Y. Nakamura for their technical assistance. This work was supported by the Core Research for Evolutional Science and Technology, Japan Science and Technology Agency and by a Grant-in-Aid from the Ministry of Health, Labor and Welfare of Japan for the third-term Comprehensive 10-year Strategy for Cancer Control.

Author Contributions Y.C., Y.L.C. and J.T. contributed equally to this work. M.K. and M.Sa. performed microarray experiments and subsequent data analyses. Y.C. and J.T. performed mutation analysis of *ALK*. Y.C., Y.L.C., J.T., M.So., L.W. and H.M. conducted functional assays of mutant ALK. A.N., M.O., T.I., A.K. and Y.H. prepared tumour specimens and were involved in statistical analysis. A.N., Y.H., H.M., J.T. and S.O. designed the overall study, and S.O. and J.T. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information The nucleotide sequences of *ALK* mutations detected in this study have been deposited in GenBank under the accession numbers EU788003 (K1062M), EU788004 (T1087I), EU788005 (F1174L; TTC/TTA), EU788006 (F1174L; TTC/TTG), EU788007 (F1174C), EU788008 (F1174V), EU788009 (F1245L) and EU788010 (R1275Q). The copy number data as well as the raw microarray data will be accessible from http://www.ncbi.nlm.nih.gov/geo/ with the accession number GSE12494. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.O. (sogawa-tky@umin.net) or Y.H. (hayashiy-tky@umin.ac.jp).

METHODS

Specimens. Primary neuroblastoma specimens were obtained during surgery or biopsy from patients who were diagnosed with neuroblastoma and admitted to a number of hospitals in Japan. In total, 215 primary neuroblastoma specimens were subjected to SNP array analysis after informed consent was obtained from the parents of each patient. The patients were staged according to the International Neuroblastoma Staging System[23]. The clinicopathological findings are summarized in Supplementary Table 1. Twenty-four neuroblastoma-derived cell lines were also analysed by SNP array analysis (Supplementary Table 2). The SCMC-N2, SCMC-N4 and SCMC-N5 cell lines were established in our laboratory[24,25]. The SJNB series of cells and the UTP-N-1[26] cell line were gifts from A. T. Look and A. Inoue, respectively. The other cell lines used were obtained from the Japanese Cancer Resource Cell Bank (http://cellbank.nibio.go.jp/).

Microarray analysis. High molecular mass DNA was isolated from tumour specimens as well as from the peripheral blood or the bone marrow as described previously[24]. The DNA was subjected to SNP array analysis using Affymetrix GeneChip Mapping 50K and/or 250K arrays (Affymetrix) according to the manufacturer's suggested protocol. The scanned array images were processed with Gene Chip Operation software (GCOS)[13], followed by SNP calls using GTYE. Genome-wide copy number measurements and loss of heterozygosity detection were performed using CNAG/AsCNAR algorithms[14], which enabled an accurate determination of allele-specific copy numbers.

Confirmation of SNP array data. FISH and/or genomic PCR analysis confirmed the results of SNP array analyses as described previously[13]. PCR primer sets were designed to amplify several adjacent fragments inside and outside of the homozygously deleted regions in tumour samples.

Mutation analysis. Mutations in the *ALK* gene were examined in 239 neuroblastoma samples, including 24 cell lines, by denaturing high-performance liquid chromatography (DHPLC) using the WAVE system (Model 4500; Transgenomic) according to the manufacturer's suggested protocol[16]. The samples showing abnormal conformations were subjected to direct sequencing analysis using an ABI PRISM 3100 Genetic Analyser (Applied Biosystems). Using direct sequencing, mutation analysis of *MYCN* was also performed in seven cases with *ALK* alterations but not *MYCN* amplification. The primer sets used in this study are listed in Supplementary Table 5.

Transforming potential of *ALK* mutants. Total RNA was extracted from SJNB-1 (wild type), SCMC-N2 (F1174L) and SCMC-N5 (K1062M) cells as described previously[26]. First-strand cDNA was synthesized from RNA using Transcriptor Reverse Transcriptase and an oligo (dT) primer (Roche Applied Science). The resulting cDNA was then amplified by PCR using the KOD-Plus-Ver.2 DNA polymerase (Toyobo) and the primers sense 5′-TCAGAAGCTTTACCAAGGACTGTTCAGAGC-3′ and antisense 5′-AATTGCGGCCGCTACTTGTCATCGTCGTCCTTGTAGTCGGGCCCAGGCTG GTTCATGC-3′, thereby introducing a HindIII site at the 5′ terminus and a NotI site and a Flag sequence at the 3′ terminus. The HindIII–NotI fragments of *ALK* cDNA were subcloned into pcDNA3 to generate expression plasmids. After resequencing to confirm that they had no other mutations, the *ALK* plasmids were used for transfection into NIH3T3 cells using Effectene Transfection Reagent (Qiagen) according to the suggested manufacturer's protocol. The transfected NIH3T3 cells were selected in 800 $\mu g\,ml^{-1}$ G418 for 2 weeks to obtain stably expressing clones.

To evaluate the phosphorylation status of ALK mutants, the cell lysates of stable clones were immunoprecipitated with antibodies to Flag (Sigma) and the resulting precipitates were subjected to western blot analysis with the antibody specific to pTyr 1604 (Cell Signaling Technology) of ALK and the generic anti-phosphotyrosine antibody (PY20). The *in vitro* kinase activity of ALK mutants was measured using a non-radioactive isotope solid-phase enzyme-linked immunosorbent assay using the Universal Tyrosine Kinase Assay kit (Takara) according to the manufacturer's suggested protocol. We also performed the *in vitro* kinase assay with the synthetic YFF peptide (Operon Biotechnologies) as described previously[18]. For anchorage-independent growth analysis, 1×10^3 stably transfected NIH3T3 cells were mixed in 0.3% agarose with 10% FBS-DMEM and plated on 0.6% agarose-coated 35-mm dishes. After culture for 28 days, the colonies of $>$0.1 mm in diameter were counted. The quantification of the colonies was from three independent experiments. To investigate the downstream signalling of ALK, western blot analysis was performed using the anti-ERK1/2, anti-phospho-ERK1/2, anti-AKT, anti-phospho-AKT, anti-STAT3 and anti-phospho-STAT3 antibodies (Cell Signaling Technology)[15].

The cDNA mutant of *ALK* was also inserted into the pMXS plasmid and the constructs were introduced into NIH3T3 cells by the calcium phosphate method as described previously[9]. The cells were then either cultured for 21 days or injected subcutaneously at six sites in three nude mice.

Inhibition of ALK through RNAi-mediated knockdown. To suppress the expression of the ALK protein, two different pairs of *ALK* siRNAs (*ALK* siRNA1 and *ALK* siRNA2) were obtained (Qiagen)[15]. The sequences were 5′-GAGUCUGGCAGUUGACUUCdTdT-3′ for *ALK* siRNA1 and 5′-GCUCCGGCGUGCCAAGCAGdTdT-3′ for *ALK* siRNA2. A siRNA, targeting a sequence in firefly (*Photinus pyralis*) luciferase mRNA (*luc* siRNA), was used as a negative control (Qiagen)[15]. The sequences of *luc* siRNA were as follow: sense 5′-CGUACGCGGAAUACUUCGAdTdT-3′ and antisense 5′-UCGAAGUAUUCCGCGUACGdTdT-3′. Gene knockdown was achieved in SK-N-SH and LAN-2 cells using HiPerFect transfection reagent following the manufacturer's suggested instructions (Qiagen). To assess the effect of ALK knockdown on cell growth, these cells were seeded in 96-well plates at a concentration of 8.0×10^3 cells per well 24 h before transfection and assayed using the Cell Counting Kit-8 (Wako).

Statistical analysis. The significance of the correlation between *MYCN* amplification and *ALK* mutation was tested according to the conventional 2×2 contingency table using Fisher's exact test. The significance of the differences in kinase activity between wild-type and mutant ALK kinases was examined by the Mann–Whitney *U*-test based on the measured percentage activity of kinases in the precipitates of the corresponding samples. The significance of the differences in colony formation between wild-type and mutant ALK kinases was also examined by the Mann–Whitney *U*-test. The size of the hazards from possible risk factors, including International Neuroblastoma Staging System stages, *MYCN* status and *ALK* mutation/amplification were estimated by Cox regression analysis assuming a proportional hazard model using Stata software. Correlation between ploidy and clinical stage was tested by nptrend test.

23. Smith, E. I., Haase, G. M., Seeger, R. C. & Brodeur, G. M. A surgical perspective on the current staging in neuroblastoma—the International Neuroblastoma Staging System proposal. *J. Pediatr. Surg.* **24**, 386–390 (1989).
24. Takita, J. *et al.* Allelotype of neuroblastoma. *Oncogene* **11**, 1829–1834 (1995).
25. Takita, J. *et al.* Absent or reduced expression of the caspase 8 gene occurs frequently in neuroblastoma, but not commonly in Ewing sarcoma or rhabdomyosarcoma. *Med. Pediatr. Oncol.* **35**, 541–543 (2000).
26. Takita, J. *et al.* Allelic imbalance on chromosome 2q and alterations of the caspase 8 gene in neuroblastoma. *Oncogene* **20**, 4424–4432 (2001).

 nature

LETTERS

Activating mutations in ALK provide a therapeutic target in neuroblastoma

Rani E. George[1]*, Takaomi Sanda[1]*, Megan Hanna[2,4]*, Stefan Fröhling[5], William Luther II[1], Jianming Zhang[3], Yebin Ahn[1], Wenjun Zhou[3], Wendy B. London[6], Patrick McGrady[6], Liquan Xue[7], Sergey Zozulya[9], Vlad E. Gregor[9], Thomas R. Webb[8], Nathanael S. Gray[3], D. Gary Gilliland[5], Lisa Diller[1], Heidi Greulich[2,4], Stephan W. Morris[7], Matthew Meyerson[2,4] & A. Thomas Look[1]

Neuroblastoma, an embryonal tumour of the peripheral sympathetic nervous system, accounts for approximately 15% of all deaths due to childhood cancer[1]. High-risk neuroblastomas are rapidly progressive; even with intensive myeloablative chemotherapy, relapse is common and almost uniformly fatal[2,3]. Here we report the detection of previously unknown mutations in the *ALK* gene, which encodes a receptor tyrosine kinase, in 8% of primary neuroblastomas. Five non-synonymous sequence variations were identified in the kinase domain of *ALK*, of which three were somatic and two were germ line. The most frequent mutation, F1174L, was also identified in three different neuroblastoma cell lines. *ALK* complementary DNAs encoding the F1174L and R1275Q variants, but not the wild-type *ALK* cDNA, transformed interleukin-3-dependent murine haematopoietic Ba/F3 cells to cytokine-independent growth. Ba/F3 cells expressing these mutations were sensitive to the small-molecule inhibitor of ALK, TAE684 (ref. 4). Furthermore, two human neuroblastoma cell lines harbouring the F1174L mutation were also sensitive to the inhibitor. Cytotoxicity was associated with increased amounts of apoptosis as measured by TdT-mediated dUTP nick end labelling (TUNEL). Short hairpin RNA (shRNA)-mediated knockdown of ALK expression in neuroblastoma cell lines with the F1174L mutation also resulted in apoptosis and impaired cell proliferation. Thus, activating alleles of the ALK receptor tyrosine kinase are present in primary neuroblastoma tumours and in established neuroblastoma cell lines, and confer sensitivity to ALK inhibition with small molecules, providing a molecular rationale for targeted therapy of this disease.

In a genome-wide analysis of primary neuroblastomas using single-nucleotide polymorphism (SNP) arrays, we noted high-level amplification of the *ALK* (anaplastic lymphoma kinase) gene[5]. To determine the frequency of this amplification, we analysed 94 tumours with amplification of the oncogene *MYCN* by fluorescence *in situ* hybridization (FISH), and documented 14 (15%) with concomitant *ALK* amplification (Supplementary Fig. 1), which was not detected in 51 tumours without *MYCN* amplification ($P = 0.0016$). None of the tumours had *ALK* rearrangements, such as those that have been found in other tumour types with *ALK* translocations[6–9].

We reasoned that in tumours without *ALK* amplification or translocation, acquired somatic mutations or germline sequence variants might contribute to oncogenicity. DNA re-sequencing of the *ALK* open reading frame in primary neuroblastomas identified five new non-synonymous sequence variations in conserved positions in the tyrosine kinase domain in 7 out of 93 samples (8%; Table 1 and Supplementary Fig.2). None of these variants were previously identified SNPs or known somatic mutations, on the basis of analysis of dbSNP and Sanger databases or by the genotyping of 270 samples derived from the International Hap Map Consortium[10].

Sequence analysis of matched normal samples from these patients showed that two of the sequence variants were germ line and three represented somatically acquired mutations (Table 1). The most common mutation, identified in 4.3% (4 out of 93) of the primary tumours, was a recurrent cytosine-to-adenine change in exon 23 that results in a phenylalanine-to-leucine substitution at codon 1174 (F1174L) within the kinase domain. Most of the patients with somatic *ALK* mutations had metastatic disease characterized by *MYCN* amplification, although one patient with the F1174L mutation had localized disease with favourable histology and unamplified *MYCN* (Supplementary Table 1). Four of the five *ALK* mutations involve residues that correspond to those affected by known activating mutations in the *EGFR* gene[11–14] (Supplementary Figs 2 and 3). The F1174 residue corresponds to V769 in EGFR, which is in a region of frequent mutation in both the *EGFR*[11] and *ERBB2* (ref. 12) genes. The F1245C mutation corresponds to L833V in EGFR, a gefitinib-resistant mutation in lung cancer (H.G. and M.M.,

Table 1 | Non-synonymous sequence variants of *ALK* in 93 patients and 30 cell lines

Patient sample	Exon	DNA	Protein	Domain	Germ line/somatic
443	22	C3452T	T1151M	Kinase	Germ line
472	23	C3522CA	F1174L	Kinase	Somatic
1034	23	C3522CA	F1174L	Kinase	Somatic
1110	23	C3522CA	F1174L	Kinase	Somatic
50	23	C3522CA	F1174L	Kinase	Somatic
50	24	G3700GA	A1234T	Kinase	Somatic
157	24	T3734TG	F1245C	Kinase	Somatic
411	25	G3824GA	R1275Q	Kinase	Germ line
Cell line	**Exon**	**DNA**	**Protein**	**Domain**	
LAN-6	20	G3271A	D1091N	Juxtamembrane	
KELLY	23	C3522CA	F1174L	Kinase	
SH-SY5Y	23	C3522CA	F1174L	Kinase	
LAN-1	23	C3522CA	F1174L	Kinase	
CHLA90	24	T3733G	F1245V	Kinase	
SMS-KCNR	25	G3824GA	R1275Q	Kinase	

Sequence numbering follows the Ensemble Transcript/Peptide ID: ENST00000389048.

[1]Department of Pediatric Oncology, [2]Department of Medical Oncology & Center for Cancer Genome Discovery, and [3]Department of Biological Chemistry and Molecular Pharmacology, Dana Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. [4]Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. [5]Division of Hematology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. [6]Children's Oncology Group Statistics and Data Center, University of Florida, Gainesville, Florida 32601, USA. [7]Departments of Pathology and Oncology, and [8]Department of Chemical Biology and Therapeutics, St. Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. [9]ChemBridge Research Laboratories Inc., San Diego, California 92127, USA.
*These authors contributed equally to this work.

unpublished observations). The R1275Q mutation is located adjacent to the homologous position of L858R in EGFR, which is the most common EGFR mutation in lung cancer[13,14].

The functional consequences of four of the mutations, T1151M, F1174L, A1234T and R1275Q, were determined by testing their abilities to transform interleukin-3 (IL-3)-dependent murine lymphoid Ba/F3 cells to cytokine-independent growth. Reductions in IL-3 concentration by 100-fold to 0.01 ng ml^{-1} resulted in a clear difference in cell proliferation, with the Ba/F3 cells expressing F1174L and R1275Q mutations having much higher cell numbers relative to those transduced with wild-type ALK or the T1151M mutation (Fig 1a). To generate IL-3-independent lines, we reduced the IL-3 concentration by half in successive passages of each transduced Ba/F3 line. After five passages, the Ba/F3 cells expressing the F1174L and the R1275Q ALK mutations, as well as NPM-ALK, were able to grow in medium completely lacking IL-3, whereas cells expressing T1151M or wild-type ALK did not survive. Moreover, when expressed in Ba/F3 cells, the F1174L allele, and to a lesser extent, the R1275Q allele, were associated with constitutive phosphorylation of ALK (Fig. 1b). In contrast, neither the T1151M nor the A1234T alleles showed ALK phosphorylation. Expression of the F1174L ALK protein in IL-3-deprived Ba/F3 cells was also associated with phosphorylation of downstream targets of ALK signalling such as STAT3 and AKT, whereas R1275Q was associated with phosphorylation of ERK1/2 and AKT (Fig. 1b). Together, these studies demonstrate that the ALK mutant proteins F1174L and R1275Q possess gain-of-function kinase activity that can sustain important signalling pathways in Ba/F3 cells cultured in the absence of IL-3.

The Ba/F3 assay has been validated for a broad spectrum of oncogenic tyrosine kinase alleles including mutant EGFR[15] and FLT3 (ref. 16), and thus we treated Ba/F3 cells expressing each of the ALK mutations with increasing concentrations of TAE684, a highly potent ALK inhibitor[4,17,18]. The activating mutation, F1174L, was found to be extremely sensitive to TAE684, with a half-maximal inhibitory concentration (IC_{50}) of 8 nM, identical to that of NPM-ALK-expressing Ba/F3 cells. The R1275Q mutation was also sensitive to TAE684, albeit with a much higher IC_{50} of 328 nM. In contrast, Ba/F3 cells expressing internal tandem duplications of the *FLT3* gene (FLT3-ITD) or wild-type ALK, did not respond to TAE684 (IC_{50} 4.6 μM; Fig. 1c).

Analysis of the *ALK* gene in a panel of 30 neuroblastoma cell lines revealed sequence variants in 6, including 3 different cell lines containing the F1174L mutation (KELLY, SH-SY5Y and LAN-1), which was also the most common mutation in the primary tumours (Table 1). An R1275Q mutation, identical to the one found in primary sample 411, was also detected in the SMS-KCNR cell line. We observed dose-dependent growth inhibition of the SH-SY5Y (F1174L) and KELLY (F1174L) neuroblastoma cell lines with increasing concentrations of TAE684 (IC_{50} values of 258 and 416 nM, respectively; Fig. 2a). These results are in agreement with data from a recent study showing sensitivity of these cell lines to TAE684 (ref. 19). Of note, the SMS-KCNR cell line expressing the ALK R1275Q mutation was resistant to TAE684 (IC_{50} of 4.9 μM; Fig. 2a), although Ba/F3 cells expressing this mutation became IL-3-independent and were sensitive to the inhibitor (Fig. 1c). Neuroblastoma cell lines without *ALK* mutations, including IMR5, were also resistant to TAE684 (Fig. 2a and Supplementary Fig. 4a). Treatment with TAE684 (200 nM) resulted in increased apoptosis in KELLY (F1174L) and SH-SY5Y (F1174L) cells, but not in the SMS-KCNR (R1275Q) and IMR5 (wild type) cells (Fig. 2b). Cytotoxicity was also associated with G1-phase arrest and substantial reductions in S-phase cell fractions (Supplementary Fig. 4b).

After treatment with TAE684 (100 nM), the sensitive cell lines SH-SY5Y (F1174L) and KELLY (F1174L) demonstrated reduced phosphorylation of ALK, ERK1/2 and AKT, and STAT3 to a lesser extent (Supplementary Fig 4c). In contrast, there was no apparent effect on the phosphorylation of AKT and STAT3 in the resistant cell line IMR5 (wild type), although there was a slight reduction in the phosphorylation of ERK1/2. Moreover, knockdown of ALK in the KELLY and SH-SY5Y cell lines (F1174L), but not the inhibitor-resistant SMS-KCNR line (R1275Q), was associated with a reduction in cell proliferation and increased apoptosis (Fig. 2c, d). In neuroblastoma, ALK is detected as both a 220 kDa protein, reflecting the glycosylated protein encoded by the transduced ALK cDNA and a second protein of ~140 kDa (Fig. 2c, inset), which has been documented by several investigators and probably represents an as yet uncharacterized splice variant-encoded isoform of ALK[20–22]. The resistance of the SMS-KCNR cell line is apparently not due to any difference between the R1275Q and F1174L mutations in ALK activation *per se*, because both of these mutations transform Ba/F3 cells to IL-3 independence and the transformed cells respond to the inhibitor (Fig. 1c). Instead, we suspect that other molecular aberrations, such as co-activation of other receptor tyrosine kinases[23,24], could have been acquired during culture that render this cell line independent of activated ALK for growth and viability.

We observed that ALK was expressed at substantially lower levels in the TAE684-sensitive neuroblastoma cell lines (KELLY (F1174L) and SH-SY5Y (F1174L)) than in the remaining cell lines harbouring either wild-type *ALK* or the R1275Q mutation (Fig. 3a). However, TAE684 inhibition of ALK kinase activity resulted in an increase in the ALK protein level in the sensitive KELLY (F1174L) cell line, but not in the IMR5 (wild type; Fig. 3b) cell line. Blockade of protein degradation by the proteasome inhibitor MG-132 resulted in increased ALK levels in KELLY (F1174L) cells, but not in IMR5 cells (Fig. 3c), consistent with a higher turnover rate in cells with the constitutively activated mutant ALK protein.

The studies reported here demonstrate previously unrecognized activating mutations affecting critical residues within the ALK kinase domain and indicate that ALK has potential as a therapeutic target in

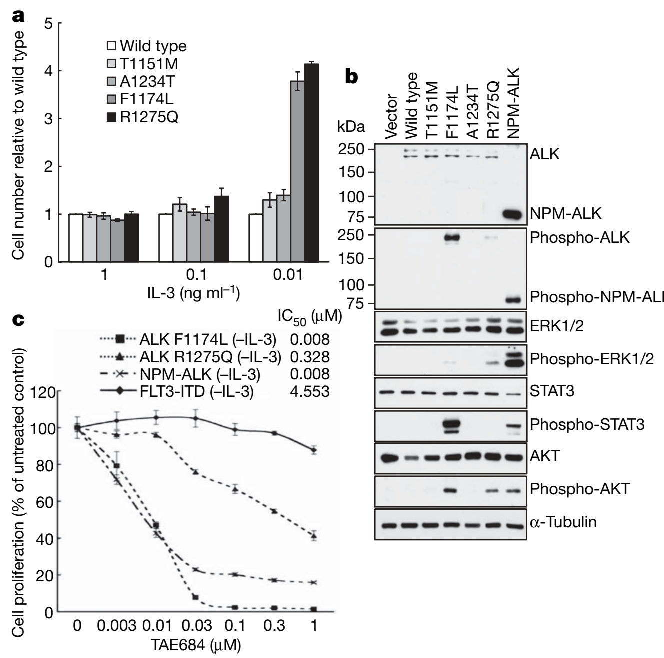


Figure 1 | *ALK* mutant alleles F1174L and R1275Q are activating in Ba/F3 cells and are sensitive to pharmacological inhibition. **a**, Growth of Ba/F3 cells expressing wild-type or mutant ALK in 10- and 100-fold-reduced concentrations of IL-3. The values are means ± s.d. of triplicate experiments. **b**, Western blot analysis of ALK proteins and their downstream effectors in wild-type or mutated ALK-expressing Ba/F3 cells depleted of IL-3 for 6 h. The mobilities of molecular mass standards (kDa) are shown on the left. **c**, Growth of mutated ALK-expressing Ba/F3 cells exposed to TAE684 for 72 h. The values are means ± s.d. of triplicate experiments.

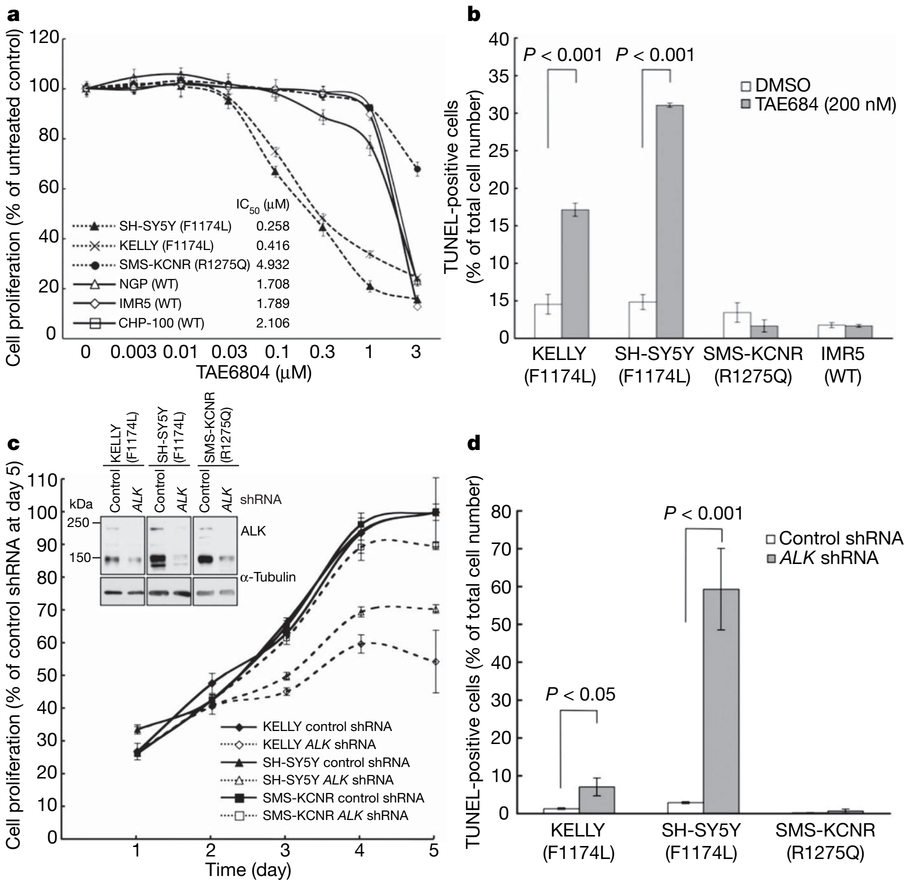


Figure 2 | Neuroblastoma cell lines harbouring the F1174L ALK mutation, but not the R1275Q ALK mutation, are dependent on the altered protein for growth and survival. **a**, Growth rates of neuroblastoma cell lines with and without ALK mutations after a 3-day exposure to varying concentrations of TAE684. The values are means ± s.d. of triplicate experiments. **b**, Induction of apoptosis in the TAE684-sensitive and resistant cell lines as determined by TUNEL assay. **c**, Growth of ALK-mutant neuroblastoma cell lines KELLY and SH-SY5Y (F1174L) and SMS-KCNR (R1275Q) in which ALK expression was downregulated using shRNA. The inset panel shows western blot analysis of ALK expression in the control and shRNA transduced lines. The mobilities of molecular mass standards (kDa) are shown on the left. **d**, Induction of apoptosis by *ALK* shRNA knockdown as determined by TUNEL assay.

neuroblastoma. Our results with the ALK small-molecule inhibitor, TAE684, demonstrate that most of the neuroblastoma cell lines harbouring activating ALK mutations are dependent on the altered ALK protein for survival. One exception is the activating R1275Q allele, which when expressed by itself in Ba/F3 cells is sensitive to treatment with TAE684, but not in the SMS-KCNR neuroblastoma background. This observation, together with the lack of transforming activity of the T1151M and the A1234T alleles, supports the emerging concept that mutations affecting critical domains of cancer genes must be studied both genetically and functionally to validate their potential as therapeutic targets[16]. Thus, it will be important to test the ability of each mutant ALK protein identified in patient tumour samples to confer IL-3 resistance in Ba/F3 cells and mediate sensitivity to ALK inhibitory drugs as they move into therapeutic trials. One sensitive cell line, KELLY (F1174L), harboured both an activating *ALK* mutation and *MYCN* amplification, suggesting that a subset of very high risk *MYCN*-amplified neuroblastomas may respond to treatment with an ALK inhibitor. Although established neuroblastoma cell lines with high levels of wild-type ALK expression did not respond to ALK inhibitors in our study, it will be important to evaluate whether the larger group of neuroblastoma patients expressing wild-type ALK proteins have responses when ALK inhibitors are tested *in vivo* in the natural tumour microenvironment, which includes exposure to the effects of ALK ligands[25,26].

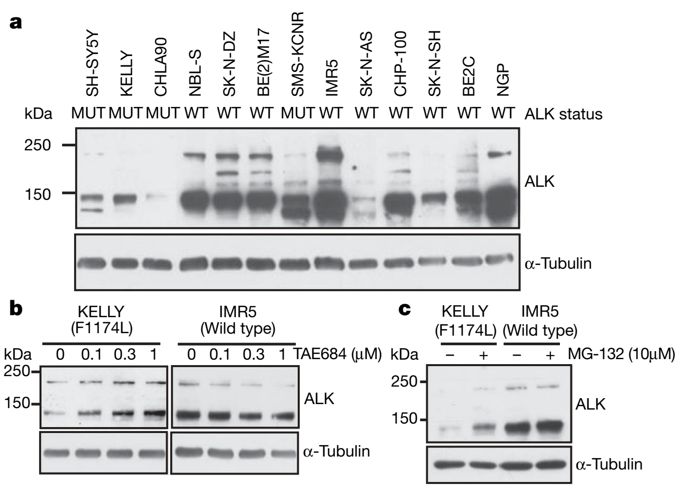


Figure 3 | The constitutively activated F1174L ALK protein has a higher turnover rate than wild-type ALK in neuroblastoma cell lines. **a**, Analysis of ALK expression in wild-type (WT) and ALK-mutated (MUT) neuroblastoma cell lines. **b**, Western blot depicting ALK expression in the mutated cell line KELLY (F1174L) and in the ALK wild-type cell line IMR5 after exposure to increasing doses of TAE684. **c**, Western blot of ALK expression in the ALK-mutated cell line KELLY (F1174L), and in the IMR5 cell line harbouring wild-type ALK, after treatment with the proteasome inhibitor MG-132. The mobilities of molecular mass standards (kDa) are shown on the left.

METHODS SUMMARY

DNA sequencing. Primers were designed to cover the 29 exons of *ALK* and characterized using 3 Coriell DNAs. Passing primers and samples were amplified by PCR and sequenced bidirectionally on an ABI 3730xl automated sequencer (Applied Biosystems). Automated analysis and coverage statistics were generated with SNP Compare (an in-house package using PolyPhred[27] and PolyDHAN (D. Richter *et al.*, manuscript in preparation)). Bidirectional sequence traces were analysed with Mutation Surveyor (Softgenetics, version 3.10) and manual review. Genotyping was performed by primer extension mass spectrometry.

Cell culture. Neuroblastoma cell lines were cultured in RPMI-1640 containing L-glutamine and 10% fetal bovine serum (FBS; Sigma-Aldrich). Ba/F3 cells were maintained in RPMI-1640 supplemented with 10% FBS and 0.5 ng ml^{-1} murine IL-3 (Millipore).

DNA constructs and retrovirus production. *ALK* mutations were engineered using the QuikChange II Site-Directed Mutagenesis Kit (Stratagene). The

mutant cDNAs were subcloned into the pMSCV-Neo-luc retroviral vector. Co-transfection of 293T cells and infection of Ba/F3 cells with retroviral supernatants were performed as described previously[16]. Transduced Ba/F3 cells were selected with G418 for 7 days and subjected to Ficoll separation to isolate surviving cells.

Cytokine independence assays. Ba/F3 cells transduced with each of the pMSCV-Neo-luc constructs were seeded at $1 \times 10^5\,ml^{-1}$ and treated with 1, 0.1 and 0.01 $ng\,ml^{-1}$ IL-3 for 72 h. The number of viable cells was determined by trypan blue exclusion using a Vi-CELL Series Cell Viability Analyzer (Beckman Coulter).

Drug sensitivity assay. Cell viability was tested 72 h after addition of the compound by CellTiter-Glo Luminescent Cell Viability Assay (Promega). IC_{50} values were calculated by nonlinear regression (variable slope) using GraphPad Prism 5 software.

Immunoblotting. Immunoblotting was performed as described previously[16]. Antibodies are listed in Supplementary Methods.

Proteasome inhibitor treatment. 1×10^6 cells were treated with the proteasome inhibitor MG-132 (Sigma-Aldrich) for 1 h, washed in PBS and immunoblotted as described previously[16].

Received 7 June; accepted 18 August 2008.

1. National Cancer Institute.. Surveillance, Epidemiology and End Results Database. <http://seer.cancer.gov> (2005).
2. Matthay, K. K. *et al.* Treatment of high-risk neuroblastoma with intensive chemotherapy, radiotherapy, autologous bone marrow transplantation, and 13-cis-retinoic acid. Children's Cancer Group. *N. Engl. J. Med.* **341,** 1165–1173 (1999).
3. George R. E. . *et al.* High risk neuroblastoma treated with tandem autologous peripheral-blood stem cell-supported transplant: long-term survival update. *J. Clin. Oncol.* **24,** 2891–2896 (2006).
4. Galkin A. & V, et a. l. Identification of NVP-TAE684, a potent, selective, and efficacious inhibitor of NPM-ALK. *Proc. Natl Acad. Sci. USA* **104,** 270–275 (2007).
5. George, R. E. *et al.* Genome-wide analysis of neuroblastomas using high-density single nucleotide polymorphism arrays. *PLoS One* **2,** e255–264 (2007).
6. Morris S.W. et al.. Fusion of a kinase gene, *ALK*, to a nucleolar protein gene, *NPM*, in non-Hodgkin's lymphoma. *Science* **263,** 1281–1284 (1994).
7. Soda, M. *et al.* Identification of the transforming *EML4–ALK* fusion gene in non-small-cell lung cancer. *Nature* **448,** 561–566 (2007).
8. Rikova K. . *et al.* Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* **131,** 1190–1203 (2007).
9. Chiarle, R., Voena, C., Ambrogio, C., Piva, R. & Inghirami, G. The anaplastic lymphoma kinase in the pathogenesis of cancer. *Nature Rev. Cancer* **8,** 11–23 (2008).
10. The International HapMap Consortium. The International HapMap Project. *Nature* **426,** 789–796 (2003).
11. Greulich, H. *et al.* Oncogenic transformation by inhibitor-sensitive and -resistant EGFR mutants. *PLoS Med.* **2,** e313 (2005).
12. Stephens, P. *et al.* Lung cancer: intragenic ERBB2 kinase mutations in tumours. *Nature* **431,** 525–526 (2004).
13. Lynch, T. J. *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **350,** 2129–2139 (2004).
14. Paez, J. G. *et al.* EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304,** 1497–1500 (2004).
15. Jiang, J. *et al.* Epidermal growth factor-independent transformation of Ba/F3 cells with cancer-derived epidermal growth factor receptor mutants induces gefitinib-sensitive cell cycle progression. *Cancer Res.* **65,** 8968–8974 (2005).
16. Fröhling, S. *et al.* Identification of driver and passenger mutations of FLT3 by high-throughput DNA sequence analysis and functional assessment of candidate alleles. *Cancer Cell* **12,** 501–513 (2007).
17. Koivunen, J. P. *et al. EML4–ALK* fusion gene and efficacy of an ALK kinase inhibitor in lung cancer. *Clin. Cancer Res.* **14,** 4275–4283 (2008).
18. Li, R. & Morris, S. W. Development of anaplastic lymphoma kinase (ALK) small-molecule inhibitors for cancer therapy. *Med. Res. Rev.* **28,** 372–412 (2008).
19. McDermott, U. *et al.* Genomic alterations of anaplastic lymphoma kinase may sensitize tumors to anaplastic lymphoma kinase inhibitors. *Cancer Res.* **68,** 3389–3395 (2008).
20. Lamant, L. *et al.* Expression of the ALK tyrosine kinase gene in neuroblastoma. *Am. J. Pathol.* **156,** 1711–1721 (2000).
21. Miyake, I. *et al.* Activation of anaplastic lymphoma kinase is responsible for hyperphosphorylation of ShcC in neuroblastoma cell lines. *Oncogene* **21,** 5823–5834 (2002).
22. Osajima-Hakomori, Y. *et al.* Biological role of anaplastic lymphoma kinase in neuroblastoma. *Am. J. Pathol.* **167,** 213–222 (2005).
23. Stommel, J. M. *et al.* Coactivation of receptor tyrosine kinases affects the response of tumor cells to targeted therapies. *Science* **318,** 287–290 (2007).
24. Engelman, J. A. *et al. MET* amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science* **316,** 1039–1043 (2007).
25. Nakagawara, A. *et al.* Differential expression of pleiotrophin and midkine in advanced neuroblastomas. *Cancer Res.* **55,** 1792–1797 (1995).
26. Calvet, L. *et al.* Pleiotrophin, a candidate gene for poor tumor vasculature and *in vivo* neuroblastoma sensitivity to irinotecan. *Oncogene* **25,** 3150–3159 (2006).
27. Nickerson, D. A., Tobe, V. O. & Taylor, S. L. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res.* **25,** 2745–2751 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Elechko and L. Moreau for technical assistance, A. Azarova for help with cell culture and cell growth assays, Q. Jiang and X. Cui for assistance during ALK small-molecule inhibitor development, and A. Kung for Ba/F3 cells and retroviral vectors. We acknowledge the Children's Oncology Group for provision of neuroblastoma tumour and matched normal DNAs, and tumour touch prep slides. We thank J. Maris for neuroblastoma cell lines. We thank Abbott Molecular International for the Vysis LSI *ALK* Dual Color, Break Apart Rearrangement Probe. This work was supported by grants from the National Institutes of Health (R.E.G.), the Friends for Life Neuroblastoma Fund (R.E.G.), the Children's Oncology Group (R.E.G.), Alex's Lemonade Stand Foundation (M.H. and M.M.), NCI CA69129 (L.X. and S.W.M.), Cancer Center Core grant CA21765 (T.R.W., L.X. and S.W.M.), the American Lebanese Syrian Associated Charities and St. Jude Children's Research Hospital (T.R.W., L.X. and S.W.M.).

Author Contributions R.E.G., M.M. and A.T.L. designed the experiments and wrote the manuscript. M.H., H.G. and M.M. performed the DNA sequencing and analysis. T.S., S.F., W.L., Y.A., H.G. and R.E.G. carried out the functional analyses. J.Z., W.Z. and N.S.G. performed the homology modelling and synthesis of TAE684. W.B.L. and P.M. performed the statistical analysis. S.Z., V.E.G. and T.R.W. were involved with the design of ALK inhibitors. L.X. and S.W.M. assisted with experimental design, provided reagents and advice on ALK inhibitors. M.M., L.D. and D.G.G. provided advice on the manuscript.

Author Information Microarray data have been submitted to the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) public database. The accession numbers for the SNP array analyses are GSM206563 and GSM206564. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.T.L. (thomas_look@dfci.harvard.edu) or M.M. (matthew_meyerson@dfci.harvard.edu).

Vol 455|16 October 2008|doi:10.1038/nature07358

The mode of Hedgehog binding to Ihog homologues is not conserved across different phyla

Jason S. McLellan[1]*, Xiaoyan Zheng[2]*, Glenn Hauk[1], Rodolfo Ghirlando[4], Philip A. Beachy[2,3] & Daniel J. Leahy[1]

Hedgehog (Hh) proteins specify tissue pattern in metazoan embryos by forming gradients that emanate from discrete sites of expression and elicit concentration-dependent cellular differentiation or proliferation responses[1,2]. Cellular responses to Hh and the movement of Hh through tissues are both precisely regulated, and abnormal Hh signalling has been implicated in human birth defects and cancer[3–7]. Hh signalling is mediated by its amino-terminal domain (HhN), which is dually lipidated and secreted as part of a multivalent lipoprotein particle[8–10]. Reception of the HhN signal is modulated by several cell-surface proteins on responding cells, including Patched (Ptc), Smoothened (Smo), Ihog (known as CDO or CDON in mammals) and the vertebrate-specific proteins Hip (also known as Hhip) and Gas1 (ref. 11). *Drosophila* Ihog and its vertebrate homologues CDO and BOC contain multiple immunoglobulin and fibronectin type III (FNIII) repeats, and the first FNIII repeat of Ihog binds *Drosophila* HhN in a heparin-dependent manner[12,13]. Surprisingly, pull-down experiments suggest that a mammalian Sonic hedgehog N-terminal domain (ShhN) binds a non-orthologous FNIII repeat of CDO[12,14]. Here we report biochemical, biophysical and X-ray structural studies of a complex between ShhN and the third FNIII repeat of CDO. We show that the ShhN–CDO interaction is completely unlike the HhN–Ihog interaction and requires calcium, which binds at a previously undetected site on ShhN. This site is conserved in nearly all Hh proteins and is a hotspot for mediating interactions between ShhN and CDO, Ptc, Hip and Gas1. Mutations in vertebrate Hh proteins causing holoprosencephaly and brachydactyly type A1 map to this calcium-binding site and disrupt interactions with these partners.

CDO is composed of five immunoglobulin domains followed by three FNIII repeats, a membrane-spanning region and a cytoplasmic tail[15]. The second FNIII repeat of CDO (CDOFn2) is homologous to the Ihog repeat that binds *Drosophila* HhN and heparin, but the third FNIII repeat (CDOFn3) seems to be most important for binding Shh[12,14] (Figs 1a and 2a). However, purified CDOFn2, CDOFn3 and a tandem repeat of these domains (CDOFn23) all failed to interact appreciably with ShhN in the presence or absence of heparin, as judged by size-exclusion chromatography (SEC). Pull-down experiments showing an interaction between ShhN and CDOFn3 were performed in the presence of serum[12,14], and a search for additional factors involved in ShhN–CDO binding led to the discovery that calcium ions promote high-affinity interactions between ShhN and CDOFn3 (Supplementary Fig. 1). Analytical ultracentrifugation and isothermal titration calorimetry (ITC) studies demonstrate that the interaction between ShhN and CDOFn3 is 1:1, requires calcium, is not heparin-dependent, and has a dissociation constant of ~1.3 μM (Figs 1a and 2c, and Supplementary Fig. 2).

The calcium-dependent ShhN–CDOFn3 complex was crystallized, and its structure determined by molecular replacement and refined to R/R_{free} of 0.186/0.229 at 1.7 Å resolution (Supplementary Table 1). The ShhN–CDOFn3 interface buries 1,325 Å^2 with a shape complementarity of 0.61 (Fig. 1b, c)—values consistent with a physiological interface[16,17]. This interface encompasses a hydrophilic region involving Lys 88, Arg 124, Arg 154, Arg 156, Ser 178 and Lys 179 on ShhN and Asp 870, Asp 872, Glu 897, Asp 901 and Glu 922 on CDOFn3 as well as a hydrophobic region involving Val 918, Met 919 and Ile 920 on CDOFn3 and His 134 and the methylene groups of Glu 90 on ShhN (Fig. 1c and Supplementary Figs 3–5). Binding studies with four previously created clusters of surface mutations in ShhN (surfaces A, B, C and D)[12] and five new mutations support identification of this interface as functionally relevant (Supplementary Fig. 3). Sequence analysis suggests that BOC, a receptor for Shh in axon guidance[18], is likely to bind ShhN in the same manner as CDO (Supplementary Figs 5 and 6).

The structure of the ShhN–CDOFn3 complex reveals a previously unknown binuclear calcium-binding site in ShhN buried at the interface with CDOFn3, which accounts for the calcium-dependence of this interaction (Fig. 1). The calcium ions are coordinated by three aspartate and three glutamate residues in ShhN (Asp 96, Asp 130, Asp 132, Glu 90, Glu 91 and Glu 127) in an arrangement similar to binuclear calcium sites in other proteins[19]. Each ion is coordinated by eight oxygen atoms with an average distance of 2.5 Å, consistent with calcium coordination[20]. Identification of these ions as calcium was confirmed by anomalous difference Fourier analysis (Supplementary Fig. 7). The level of conservation of the six calcium-coordinating residues in Hh sequences suggests that calcium binding is a general property of Hh proteins (Supplementary Fig. 8). Attempts to measure the dissociation constant of ShhN for calcium indicate a K_d of >100 μM. This weak affinity explains the failure to observe this calcium site in earlier HhN crystal structures, as calcium would have dissociated during purification[13,21]. One millimolar Ca^{2+} is sufficient to restore calcium-dependent activity to ShhN *in vitro*, and the presence of ~1 mM calcium ions in the extracellular space suggests that calcium is likely to be a constitutive component of HhN proteins *in vivo*.

Notably, the ShhN–CDOFn3 interaction is completely unlike the interaction between *Drosophila* HhN and Ihog[13] (Fig. 2b). The surfaces on ShhN and HhN that bind CDO and Ihog, respectively, overlap at only a single residue and interact with non-orthologous domains of CDO and Ihog (CDOFn3 versus IhogFn1). This result is surprising given the 68% sequence identity between human SHHN and *Drosophila* HhN and sequence and functional data implicating CDO and BOC as the closest vertebrate homologues of Ihog[12,14,22]. To confirm the distinct binding modes, four simultaneous amino acid

[1]Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [2]Department of Developmental Biology, and [3]Howard Hughes Medical Institute Stanford University School of Medicine, Stanford, California 94305, USA. [4]Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Disease, National Institutes of Health, Bethesda, Maryland 20892, USA.
*These authors contributed equally to this work.

substitutions (T41Y, A44V, Y45L and K179R) were introduced into ShhN to create a variant, ShhN-Fly, that contains all Ihog-contacting residues found at the *Drosophila* HhN–Ihog interface[13]. ShhN-Fly binds Ihog with 100-fold greater affinity than wild-type ShhN and has a slightly stronger affinity for Ihog (K_d of ~0.5 μM) than does *Drosophila* HhN (K_d of ~2.6 μM; Fig. 2c, d). ShhN-Fly does not bind well to cells expressing *Drosophila* Ptc (Supplementary Fig. 9). As is the case for HhN, ShhN-Fly requires heparin but not calcium to bind Ihog, and the interaction is exothermic (Fig. 2c, d). In contrast, the interaction between ShhN or ShhN-Fly and CDO is endothermic (Fig. 2c and Supplementary Fig. 10). Expression of Shh-Fly within clones in the anterior compartment of *Drosophila* wing imaginal discs also induces expression of the Hh pathway target gene *ptc* to a greater extent than wild-type Shh, and this expression extends several cells beyond the clone boundary (Supplementary Fig. 11).

Although modelling suggests that a complex containing ShhN-Fly, IhogFn12 and CDOFn3 is possible, this complex is not observed by SEC. The CDO-binding site on ShhN overlaps extensively with a heparin-binding site required for interactions between HhN and Ihog[13], and heparin may impede binding of CDOFn3 to ShhN-Fly. Analysis of Hh sequences suggests that the CDO-like binding mode is conserved in vertebrates and the Ihog-like mode in arthropods, but the binding mode for intermediate and earlier branchpoint clades is not clear.

Given the importance of this previously unappreciated calcium-binding site for Shh–CDO interactions, we examined its role in mediating interactions between ShhN and other ShhN-binding partners. Four simultaneous amino-acid substitutions in the ShhN calcium-binding site (E90Q, E91Q, E127Q and D132N) or an increase in concentration of EGTA, a calcium chelator, led to a >10-fold reduction in ShhN binding to Hip-transfected cells and a ~2-fold reduction in binding to Ptc-transfected cells when compared to wild-type ShhN (Fig. 3a, b). EGTA also reduces binding of *Drosophila* HhN to *Drosophila* Ptc-transfected cells (Supplementary Fig. 12). The effects of EGTA seem stronger, possibly owing to effects of calcium chelation on the cell membrane or chelation of a previously identified zinc ion in ShhN[21]. The presence of increasing amounts of soluble CDOFn3 or CDOFn23, but not CDOFn2, also reduced the ability of ShhN to bind to Hip- and Ptc-transfected cells of human or fly origin (Fig. 3c and Supplementary Figs 13 and 14), suggesting that Hip and Ptc may compete directly with CDO for binding to ShhN. Although consistent with our results, direct competition would conflict with an earlier observation of synergy between CDO and Ptc for Shh binding[23]. Resolution of this issue awaits studies with purified proteins.

Gas1 has also been reported to interact with Shh[23], and a pull-down assay was developed in which Fc-tagged Gas1 immobilized on protein G beads was used to precipitate ShhN. The absence of calcium in the precipitating medium or the presence of CDOFn3 (but not CDOFn2) greatly reduced the ability of Gas1 to precipitate ShhN (Fig. 3d, e). These results suggest that an intact and accessible calcium-binding site on ShhN is required to mediate interactions between ShhN and Gas1.

A missense substitution in human SHH (D88V) that causes holoprosencephaly (HPE)[24], a developmental disorder involving abnormal formation of midline face and forebrain features, maps to a loop on ShhN that contains three of the six calcium-coordinating residues (Fig. 4a). Asp 88 does not directly contact CDOFn3 or coordinate calcium but stabilizes the conformation of the calcium-coordinating loop by forming hydrogen bonds to two backbone amides.

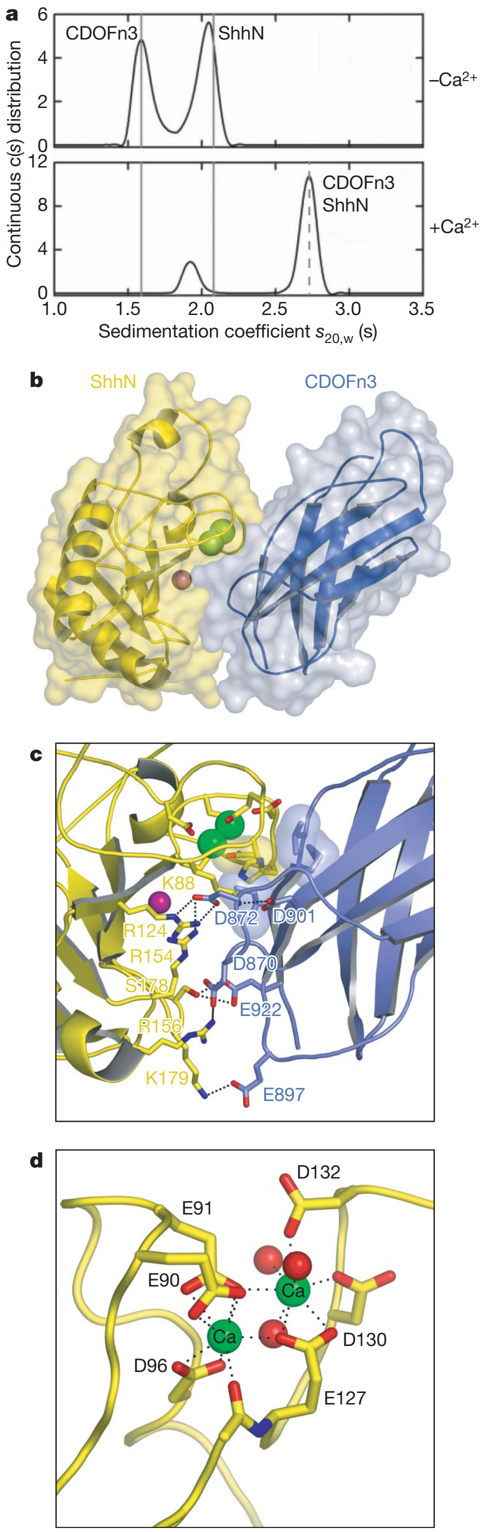


Figure 1 | A calcium-binding site on Shh is important for interactions with CDOFn3. **a**, Sedimentation velocity results for a 1:1 mixture of 20.5 μM ShhN and CDOFn3 in the absence (top) and presence (bottom) of 1 mM $CaCl_2$. Solid grey lines indicate the experimental sedimentation coefficients for monomeric CDOFn3 and ShhN. The dashed line indicates the sedimentation coefficient calculated in HYDROPRO for the 1:1 complex. **b**, Semi-transparent molecular surface over a ribbon diagram of ShhN (yellow) bound to CDOFn3 (blue). Calcium and zinc ions are depicted as green and purple spheres, respectively. **c**, The ShhN–CDOFn3 interface shown in the same orientation as in **b**. The mostly acidic residues on CDOFn3 that interact with mostly basic residues on ShhN are labelled. Semi-transparent surface representations are shown for CDO Val 918, Met 919 and Ile 920, which make van der Waals contacts with Shh His 134 and the methylene groups of Glu 90. **d**, Side chains in ShhN directly coordinating the two calcium ions are shown in stick representation. Calcium ions are green and three water molecules are shown as red spheres. Hydrogen bonds are depicted as dotted lines. All structure images were rendered with PyMOL (http://www.pymol.org).

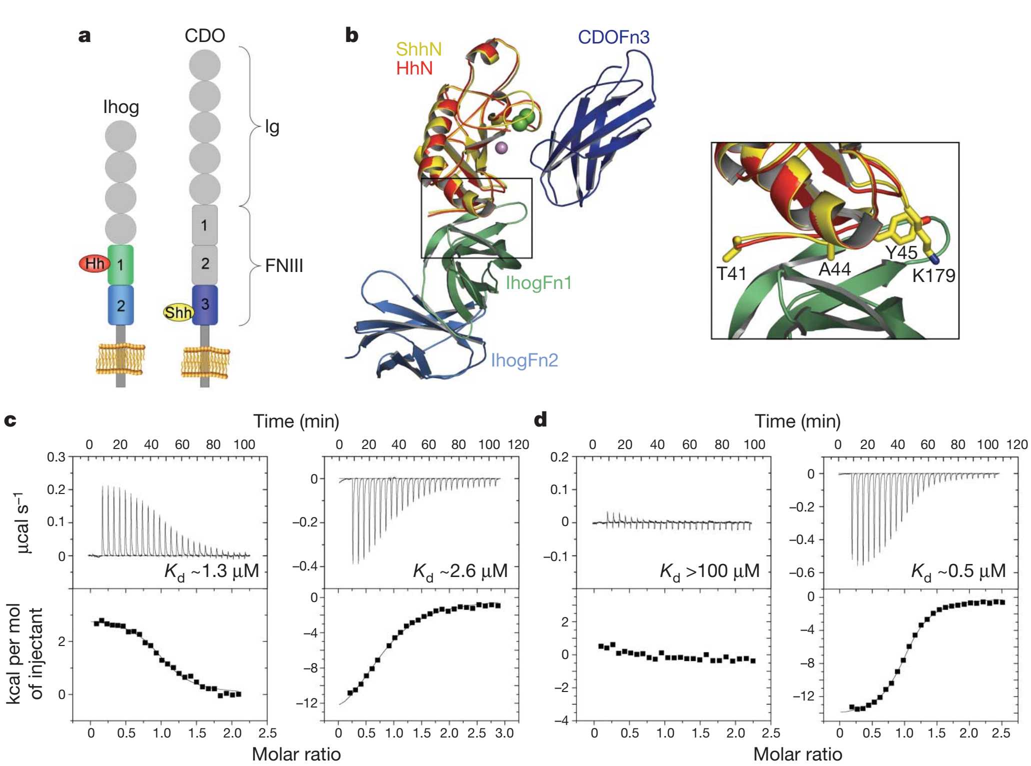


Figure 2 | Shh binds CDO differently than Hh binds Ihog. **a**, Schematic of Ihog and CDO. **b**, Ribbon diagram of the ShhN/CDOFn3 structure superimposed on the HhN/IhogFn1-2 structure by alignment of the Hh molecules. ShhN is coloured yellow, CDOFn3 dark blue, HhN red, IhogFn1 light green, and IhogFn2 light blue. Calcium and zinc ions are depicted as green and purple spheres, respectively. The boxed area shows a close-up of the interface between IhogFn1 and the Hh proteins. The side chains of four ShhN residues changed in the ShhN-Fly mutant are shown as sticks. **c**, ITC data for ShhN and CDOFn3 with calcium (left) and HhN and IhogFn1-2 with heparin and calcium (right). **d**, ITC data for ShhN and IhogFn1-2 with heparin and calcium (left) and ShhN-Fly and IhogFn1-2 with heparin and EGTA (right).

Introduction of the D88V substitution into ShhN results in loss of detectable binding to CDOFn3 (Fig. 4b) and reduces binding of ShhN to Hip-expressing cells by ~4-fold and to Ptc-expressing cells by ~5-fold (Fig. 4d). ShhN(D88V) also binds less well to Gas1 (Supplementary Fig. 15). Although no molecular explanation for the effects of the D88V mutation was previously apparent, it is now clear that the D88V reduces the affinity of ShhN for multiple partners through disruption of the calcium-coordinating region.

Brachydactyly type A1 (BDA1) is characterized by absent or shortened middle phalanges[5]. Five missense substitutions in human Indian hedgehog (IHH) that cause BDA1 (E95K, E95G, D100N, D100E and E131K) affect residues that directly coordinate calcium (Fig. 4a). ITC experiments show that wild-type IHH binds CDOFn3 with a K_d of ~2.7 μM in the presence of calcium (Supplementary Fig. 16), similar to the value of 1.3 μM for ShhN binding to CDOFn3. Introduction of the E131K mutation into IHHN results in loss of detectable CDOFn3 binding (Fig. 4c), and the E131K and D100E substitutions reduce binding of IHH to cells expressing Ptc by ~4-fold and 10-fold, respectively (Fig. 4e). The E131K and D100E substitutions also reduced IHH binding to Hip-expressing cells >20-fold (Fig. 4e); the presence of EGTA decreased binding of wild-type IHH to Ptc- and Hip-expressing cells by similar amounts (Supplementary Fig. 17), implicating disruption of calcium-dependent interactions as a common mechanism for BDA1-causing substitutions. The E127K substitution in Shh (equivalent to E131K in human IHH) reduced the binding of ShhN to Gas1 in pull-down experiments (Supplementary

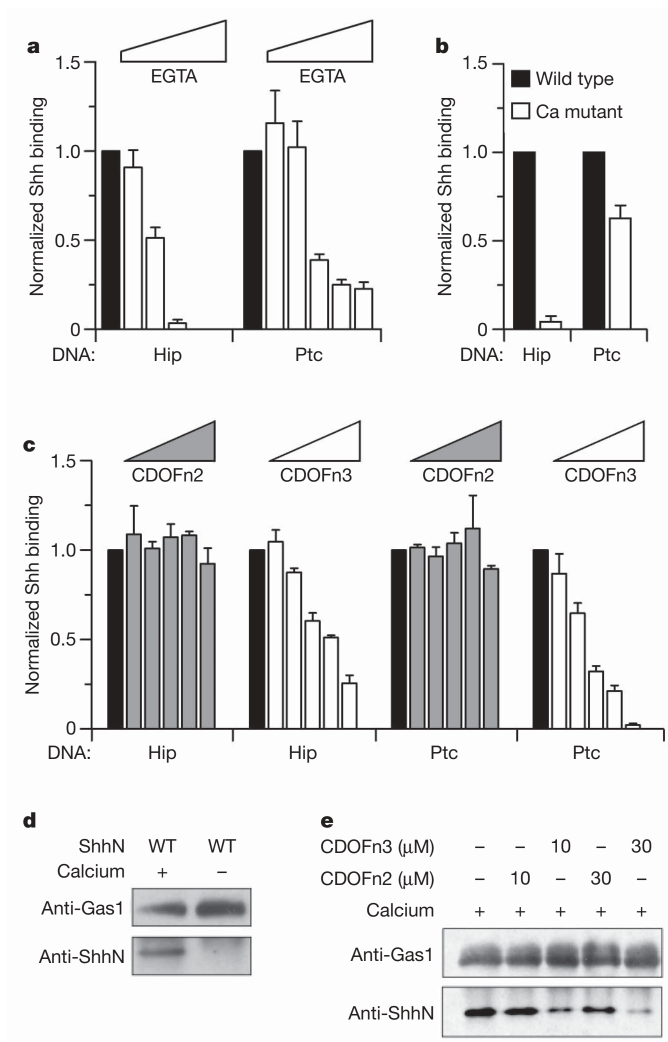


Figure 3 | The Shh calcium-binding site mediates multiple interactions. **a**, Cells expressing Hip or Ptc were assayed for Shh binding in the presence of 1.8 mM Ca^{2+} (black bars) and increasing EGTA concentrations (1.8–2.7 mM, white bars). **b**, Cell-based binding assays were performed with wild-type Shh and a Shh calcium-binding-site mutant (Ca mutant) in the presence of calcium. **c**, Cell-based binding assays with wild-type Shh were performed in the presence of calcium (black bars) and 0.5–30 μM soluble CDOFn2 (grey bars) or CDOFn3 (white bars). Error bars in **a**–**c** indicate standard deviation, with $n = 3$. **d**, Gas1 Fc-fusion protein was immobilized to protein G beads, and binding to ShhN was assayed by a western blot in the presence and absence of calcium. **e**, Gas1 pull-down experiments were repeated in the presence of CDOFn3 or CDOFn2.

Fig. 15), and substitutions in Shh equivalent to each of the five IHH BDA mutations reduced binding of Shh to Hip and Ptc in cell-based assays (Supplementary Fig. 18).

Ptc, CDO and Gas1 are positive regulators of Hh signalling[25], and Hip is a negative regulator[26]. The BDA1-causing *Ihh* mutations act in a dominant, gain-of-function manner[3], and produce a phenotype resembling a dominant mutation that produces ectopic expression of the *Shh* gene within the developing digits[27]. Among the complex effects of BDA1 mutations, disruption of interactions with Hip and an increased range of signalling seem most likely to cause a gain-of-function phenotype. Indeed, introduction of the E95K substitution into mouse Ihh (Glu 91 in mouse Shh) increases the range of Ihh action within the digit primordia of the limb, albeit with a reduced signalling potency (D. Chan *et al.*, personal communication). In contrast, the loss-of-function phenotype of the D88V mutation in Shh suggests that effects of this mutation on interactions with Ptc, CDO and/or Gas1 predominate. In general, the phenotype produced by calcium-site mutations is likely to be a composite of positive and negative effects depending on the importance or expression of various partners in affected tissues and processes.

We have shown that vertebrate Hh proteins use a previously unappreciated binuclear calcium-binding site to mediate interactions with multiple partners. Mutations causing the developmental disorders HPE and BDA1 map to this region in Shh and Ihh, respectively. Surprisingly, structural and biochemical studies demonstrate that vertebrate Hh proteins interact with the adhesion-like molecule CDO in a completely different manner than the homologous HhN–Ihog pair in *Drosophila*[13]. This difference may have arisen through convergent evolution, in which vertebrate and invertebrate ancestors independently evolved the use of homologous molecules to modulate Hh signalling. We find this scenario unlikely and suggest the following alternative. Hh proteins are secreted as part of a multivalent lipoprotein complex[9,10], which must result in high-density clustering of Hh binding partners at a localized patch on the surface of target cells. At sites of high local concentration and reduced dimensionality, weak interactions may become functionally relevant, and single mutations may have profound positive or negative effects[28]. We propose that, after evolution of an Ihog/CDO precursor that mediated Hh binding, clustering may have led to a bimodal interaction between Hh and Ihog/CDO. Additional substitutions could have resulted in the loss of one of the two binding modes in different species. Once different binding modes that supply the same function emerged, evolution of different regulatory mechanisms became possible.

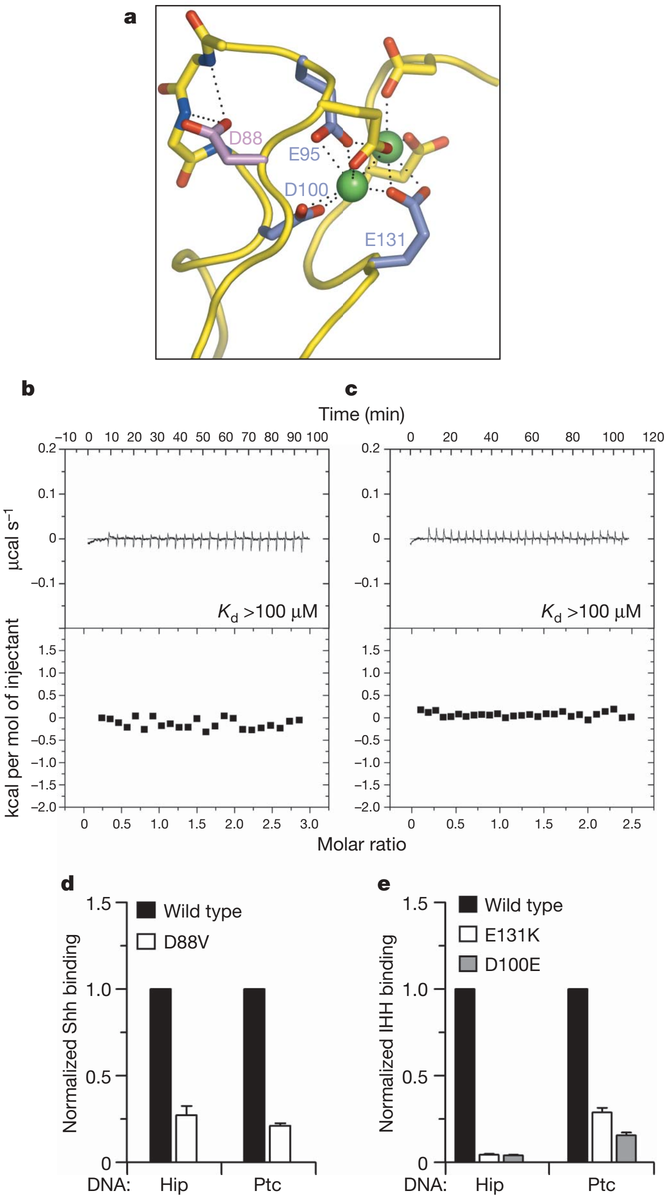


Figure 4 | HPE and BDA1 mutations in Shh and Ihh disrupt binding interactions. **a**, Loop representation of the mouse ShhN calcium-binding site. Mutation of human IHH residues E95, D100 and E131 (blue) cause BDA1. Mutation of D88 (purple) in human SHH causes HPE. Calcium ions are depicted as green spheres and hydrogen bonds are shown as dotted lines. **b**, **c**, ITC data showing loss of CDOFn3 binding to Shh(D88V) (**b**) and IHH(E131K) (**c**) in the presence of calcium. **d**, Normalized binding of Shh and Shh(D88V) to cells expressing Ptc and Hip. Error bars indicate standard deviation, with $n = 5$. **e**, Normalized binding of IHH, IHH(E131K) and IHH(D100E) to cells expressing Ptc and Hip. Error bars indicate standard deviation with $n = 3$.

METHODS SUMMARY

Sedimentation velocity experiments. Experiments were conducted at 4 °C on a Beckman Coulter XL-I analytical ultracentrifuge using multiple loading concentrations and rotor speeds of 55,000 or 60,000 r.p.m. (~244,000–290,000*g*). Scans were analysed in SEDFIT 11.3 to obtain a sedimentation coefficient, *s* or a c(*s*) distribution[29]. Values were corrected to 20 °C and water to obtain $s_{20,w}$ or c($s_{20,w}$) using solvent parameters calculated in SEDNTERP 1.09.

Crystallization. Mouse ShhN (residues 26–189) and human CDOFn3 (826–924) were mixed in the presence of 2 mM $CaCl_2$, incubated for 1 h at 22–24 °C, and the ShhN–CDOFn3 complex was purified by SEC and concentrated to ~6 mg ml^{-1} total protein. The complex was crystallized by vapour diffusion in hanging drops at 20 °C by mixing 1 μl of protein solution with 1 μl of reservoir buffer (28% PEG 3350, 200 mM $MgCl_2$, 100 mM Tris, pH 8.3). The complex crystallized in space group $C222_1$ with unit cell dimensions $a = 71.2$, $b = 98.6$ and $c = 144.0$ Å.

ITC experiments. Titrations were performed using a VP-ITC MicroCalorimeter (MicroCal Inc.). When IhogFn1–2 binding was assayed[13], 300 μM LMW heparin (Sigma) was included. CDO or Ihog proteins (143–432 μM) were added to a 1.38-ml cell containing a specific HhN protein (10–38 μM) while stirring at 260 r.p.m. Data were processed with Origin software and best fit by a single binding site model.

Cell-based binding assays. Cos1 cells plated in 6-well plates were transfected with Fugene 6 (Roche) for 48 h followed by incubation with conditioned medium containing ShhN–Ren or IhhN–Ren proteins on ice for 1 h. Cells were washed three times with cold PBS or DMEM, lysed, and 50 μl of lysate was used to measure luciferase activities.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 May; accepted 21 August 2008.
Published online 14 September 2008.

1. Ingham, P. W. & McMahon, A. P. Hedgehog signaling in animal development: paradigms and principles. *Genes Dev.* **15,** 3059–3087 (2001).
2. Hooper, J. E. & Scott, M. P. Communicating with Hedgehogs. *Nature Rev. Mol. Cell Biol.* **6,** 306–317 (2005).

3. Belloni, E. *et al.* Identification of Sonic hedgehog as a candidate gene responsible for holoprosencephaly. *Nature Genet.* **14,** 353–356 (1996).
4. Roessler, E. *et al.* Mutations in the human Sonic Hedgehog gene cause holoprosencephaly. *Nature Genet.* **14,** 357–360 (1996).
5. Gao, B. & He, L. Answering a century old riddle: brachydactyly type A1. *Cell Res.* **14,** 179–187 (2004).
6. Hahn, H. *et al.* Mutations of the human homolog of *Drosophila patched* in the nevoid basal cell carcinoma syndrome. *Cell* **85,** 841–851 (1996).
7. Johnson, R. L. *et al.* Human homolog of *patched*, a candidate gene for the basal cell nevus syndrome. *Science* **272,** 1668–1671 (1996).
8. Mann, R. K. & Beachy, P. A. Novel lipid modifications of secreted protein signals. *Annu. Rev. Biochem.* **73,** 891–923 (2004).
9. Panakova, D., Sprong, H., Marois, E., Thiele, C. & Eaton, S. Lipoprotein particles are required for Hedgehog and Wingless signalling. *Nature* **435,** 58–65 (2005).
10. Chen, M.-H., Li, Y.-J., Kawakami, T., Xu, S.-M. & Chuang, P.-T. Palmitoylation is required for the production of a soluble multimeric Hedgehog protein complex and long-range signaling in vertebrates. *Genes Dev.* **18,** 641–659 (2004).
11. Kang, J. S., Zhang, W. & Krauss, R. S. Hedgehog signaling: cooking with Gas1. *Sci. STKE* **403,** pe50 (2007).
12. Yao, S., Lum, L. & Beachy, P. The Ihog cell-surface proteins bind Hedgehog and mediate pathway activation. *Cell* **125,** 343–357 (2006).
13. McLellan, J. S. *et al.* Structure of a heparin-dependent complex of Hedgehog and Ihog. *Proc. Natl Acad. Sci. USA* **103,** 17208–17213 (2006).
14. Tenzen, T. *et al.* The cell surface membrane proteins Cdo and Boc are components and targets of the Hedgehog signaling pathway and feedback network in mice. *Dev. Cell* **10,** 647–656 (2006).
15. Kang, J.-S. *et al.* CDO: an oncogene-, serum-, and anchorage-regulated member of the Ig/fibronectin type III repeat family. *J. Cell Biol.* **138,** 203–213 (1997).
16. Conte, L. L., Chothia, C. & Janin, J. The atomic structure of protein–protein recognition sites. *J. Mol. Biol.* **285,** 2177–2198 (1999).
17. Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234,** 946–950 (1993).
18. Okada, A. *et al.* Boc is a receptor for sonic hedgehog in the guidance of commissural axons. *Nature* **444,** 369–373 (2006).
19. Chimento, D. P., Mohanty, A. K., Kadner, R. J. & Wiener, M. C. Substrate-induced transmembrane signaling in the cobalamin transporter BtuB. *Nature Struct. Mol. Biol.* **10,** 394–401 (2003).
20. Harding, M. M. The geometry of metal–ligand interactions relevant to proteins. *Acta Crystallogr. D* **55,** 1432–1443 (1999).
21. Hall, T. M. T., Porter, J. A., Beachy, P. A. & Leahy, D. J. A potential catalytic site revealed by the 1.7-A crystal structure of the amino-terminal signalling domain of Sonic hedgehog. *Nature* **378,** 212–216 (1995).
22. Kang, J.-S., Mulieri, P. J., Hu, Y., Taliana, L. & Krauss, R. S. BOC, an Ig superfamily member, associates with CDO to positively regulate myogenic differentiation. *EMBO J.* **21,** 114–124 (2002).
23. Martinelli, D. C. & Fan, C.-M. Gas1 extends the range of Hedgehog action by facilitating its signaling. *Genes Dev.* **21,** 1231–1243 (2007).
24. Heussler, H. S., Suri, M., Young, I. D. & Muenke, M. Extreme variability of expression of a Sonic Hedgehog mutation: attention difficulties and holoprosencephaly. *Arch. Dis. Child.* **86,** 293–296 (2002).
25. Allen, B. L., Tenzen, T. & McMahon, A. P. The Hedgehog-binding proteins Gas1 and Cdo cooperate to positively regulate Shh signaling during mouse development. *Genes Dev.* **21,** 1244–1257 (2007).
26. Chuang, P.-T. & McMahon, A. P. Vertebrate Hedgehog signalling modulated by induction of a Hedgehog-binding protein. *Nature* **397,** 617–621 (1999).
27. Niedermaier, M. *et al.* An inversion involving the mouse Shh locus results in brachydactyly through dysregulation of Shh expression. *J. Clin. Invest.* **115,** 900–909 (2005).
28. Kuriyan, J. & Eisenberg, D. The origin of protein interactions and allostery in colocalization. *Nature* **450,** 983–990 (2007).
29. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78,** 1606–1619 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Abramowitz and J. Schwanof of beamline X4C at the National Synchroton Light Source for assistance with X-ray data collection, C. W. Vander Kooi for suggesting ions may promote Shh–CDO interactions, W. Yang, J. Nathans, W. I. Weis, K. C. Garcia, P. A. Cole and L. M. Amzel for comments on the manuscript, and D. Chan and K. Cheah for communicating results before publication. We thank A. P. McMahon and C. M. Fan for the mouse *Hip1* and mouse *Gas1*–Fc complementary DNAs, respectively. This research was supported in part by the Intramural Research Program of the NIH, National Institute of Diabetes and Digestive and Kidney Diseases (R.G.). D.J.L. is supported by R01 HD055545 and P.A.B. is an HHMI investigator. J.S.M. is supported by a National Science Foundation Graduate Research Fellowship. X.Z. is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (DRG-1915-06).

Author Information Atomic coordinates for the Shh–CDOFn3 complex have been deposited in the Protein Data Bank with accession number 3D1M. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.J.L. (dleahy@jhmi.edu).

METHODS

Cloning, protein expression and purification. DNA fragments encoding human CDOFn2-3 (amino acids 722–924), human CDOFn3 (826–924), human BOCFn3 (710–817), *D. melanogaster* IhogFn1-2 (466–679) and IhogFn1 (466–577) were PCR amplified and cloned into the bacterial expression plasmid pT7HMT[30]. DNA fragments encoding mouse ShhN (26–189), mouse IhhN (67–231) and *D. melanogaster* HhN (86–248) were PCR amplified and cloned into a modified pMAL-c2X (NEB) bacterial expression vector. Megaprimer PCR mutagenesis was used to create the mouse Shh(D89V) (human SHH(D88V)) mutant by introducing the desired base substitutions during two rounds of PCR and subcloning the product into the expression plasmid. A similar approach was used to create the ShhN-Fly (T41Y, A44V, Y45L, K179R) and Ihh(E131K) mutants.

Proteins were expressed in BL21(DE3) cells during an overnight (16–18 h) induction. After cell lysis, proteins were partially purified using Ni-NTA resin (Qiagen) and then digested with either TEV or HRV14-3C protease to remove N-terminal tags. Proteins were then further purified by ion-exchange and size-exclusion chromatography and stored frozen at −80 °C in 20 mM Tris-HCl, pH 8.0, 200 mM NaCl.

Expression in *Drosophila* and mammalian cell cultures was driven by the actin 5C promoter and CMV promoter, respectively. *Renilla* luciferase-tagged HhN or ShhN expression vectors, and mouse Ptch1 expression vectors, have been described previously[12,31].

Data collection and structure determination. Native ShhN–CDOFn3 crystals in reservoir solution were frozen directly in the cryostream. Diffraction data were collected at beamline X4C of the National Synchrotron Light Source at Brookhaven National Laboratory and processed with HKL2000 (ref. 32). A molecular replacement solution was found with MOLREP[33] using the mouse ShhN[20] and IhogFn2 (ref. 13) structures as search models. Model building was performed using COOT[34], and refinement was carried out by CNS[35] and REFMAC[33]. Data collection and refinement statistics are presented in Supplementary Table 1. There are two complexes in the asymmetric unit and the final model consists of 3,911 protein atoms, 410 water molecules, 4 calcium ions and 2 previously described zinc ions[20].

Sedimentation equilibrium. Experiments were conducted at 4 °C on a Beckman Optima XL-A analytical ultracentrifuge. Each sample (loading volume of 130 μl) in 20 mM Tris, pH 8.0, and 200 mM NaCl was studied at three loading concentrations corresponding to a loading A_{280} of approximately 0.4, 0.8 and 1.2, and experiments were carried out at rotor speeds of 20,000 to 36,000 r.p.m. (~32,000–104,000*g*). Data were acquired as an average of four absorbance measurements at a wavelength of 280 nm using a radial spacing of 0.001 cm. Sedimentation equilibrium was achieved within 48 h and data were analysed globally in SEDPHAT 4.3 in terms of various models[36]. Solution densities ρ were calculated using SEDNTERP 1.09, as were values for the partial specific volumes v of the proteins studied. Data collected for each of the ShhN and CDOFn3 samples at different speeds and different loading concentrations were best fit in terms of a single ideal solute, returning molecular masses consistent with monomeric, monodisperse species (Supplementary Table 2). 2:1, 1:1 and 1:2 loading mixtures of ShhN and CDOFn3 containing 1 mM $CaCl_2$, each at three different values of loading A_{280}, were also studied and data were analysed in terms of a reversible $A + B \leftrightarrow AB$ or $A + A + B \leftrightarrow AAB$ equilibrium using mass conservation. Analysis in terms of an $A + B \leftrightarrow AB$ model returned excellent fits and an averaged K_d of 4 ± 2 μM. 1:1 loading mixtures of ShhN and CDOFn3 in the absence of calcium chloride, as well as Shh(D88V) and CDOFn3 in the presence of calcium chloride, were studied in a similar manner. In both of these instances no evidence for an interaction was noted and data were best fit in terms of free ShhN and CDOFn3.

Sedimentation velocity and molecular modelling. Structural data for CDOFn3, ShhN and their 1:1 complex were used to calculate the sedimentation coefficient based on a shell model with the program HYDROPRO[37] (Supplementary Table 2). Sedimentation velocity c(*s*) analyses were carried out using an *s*-value range of 0.05 to 4.0 with a linear resolution of 100 and a confidence level (*F*-ratio) of 0.68; root mean squared deviation values for the best fits ranged from 0.0035 to 0.0065 absorbance units.

Pull-down binding assays. Mouse Gas1–human-Fc fusion proteins were expressed in Cos1 cells and bound to protein G beads overnight at 4 °C. The protein G beads were then incubated with either wild-type or mutant ShhN-conditioned media overnight at 4 °C and Shh proteins associated with the matrices were detected by western blotting.

Western blot analysis. Primary antibodies were rabbit anti-Shh (1:1,000) and rabbit anti-mouse Gas1 (1:1,000)[38].

Generation of Shh- and Shh-Fly-expressing clones. Flies carrying *hsp70–FLP, UAS–GFP* were crossed to *actP* < *y* < *Gal4; UAS-Shh* (or *Shh-Fly*). 2–3 instar larvae were heat-shocked for 5 min at 37 °C to induce clones.

30. Geisbrecht, B. V., Bouyain, S. & Pop, M. An optimized system for expression and purification of secreted bacterial proteins. *Protein Expres. Purif.* **46,** 23–32 (2006).
31. Ma, Y. *et al.* Hedgehog-mediated patterning of the mammalian embryo requires transporter-like function of dispatched. *Cell* **111,** 63–75 (2002).
32. Otwinowski, Z. & Minor, W. in *Methods in Enzymology* (eds Carter, C. W. Jr & Sweet, R. M.) 307–326 (Academic, 1997).
33. The Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
34. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
35. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
36. Lebowitz, J., Lewis, M. S. & Schuck, P. Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Sci.* **11,** 2067–2079 (2002).
37. Garcia de la Torre, J., Huertas, M. L. & Carrasco, B. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* **78,** 719–730 (2000).
38. Del Sal, G., Ruaro, M. E., Philipson, L. & Schneider, C. The growth arrest-specific gene, *gas1*, is involved in growth suppression. *Cell* **70,** 595–607 (1992).

LETTERS

A role for the two-helix finger of the SecA ATPase in protein translocation

Karl J. Erlandson[1], Stephanie B. M. Miller[1], Yunsun Nam[1], Andrew R. Osborne[1]†, Jochen Zimmer[1] & Tom A. Rapoport[1]

An important step in the biosynthesis of many proteins is their partial or complete translocation across the plasma membrane in prokaryotes or the endoplasmic reticulum membrane in eukaryotes[1]. In bacteria, secretory proteins are generally translocated after completion of their synthesis by the interaction of the cytoplasmic ATPase SecA and a protein-conducting channel formed by the SecY complex[2]. How SecA moves substrates through the SecY channel is unclear. However, a recent structure of a SecA–SecY complex raises the possibility that the polypeptide chain is moved by a two-helix finger domain of SecA that is inserted into the cytoplasmic opening of the SecY channel[3]. Here we have used disulphide-bridge crosslinking to show that the loop at the tip of the two-helix finger of *Escherichia coli* SecA interacts with a polypeptide chain right at the entrance into the SecY pore. Mutagenesis demonstrates that a tyrosine in the loop is particularly important for translocation, but can be replaced by some other bulky, hydrophobic residues. We propose that the two-helix finger of SecA moves a polypeptide chain into the SecY channel with the tyrosine providing the major contact with the substrate, a mechanism analogous to that suggested for hexameric, protein-translocating ATPases.

SecA uses the energy of ATP hydrolysis to 'push' polypeptides through the SecY channel[4]. The channel has an hourglass-shaped pore that consists of funnels on the cytoplasmic and external sides of the membrane[5,6]. The constriction of the pore is located near the middle of the membrane and is formed from a 'pore ring' of hydrophobic amino acids that project their side chains radially inward. A short helix plugs the external funnel in the closed state of the channel[5] and is displaced during translocation[7,8]. The channel-interacting SecA ATPase contains two RecA-like nucleotide-binding domains (NBD1 and NBD2) that bind the nucleotide at their interface and move relative to one another during the ATP hydrolysis cycle. In addition, SecA contains a polypeptide-crosslinking domain (PPXD), a helical wing domain (HWD), and a helical scaffold domain (HSD)[9]. In a recent structure of the SecA–SecY complex[3], two helices of SecA's HSD form a 'two-helix finger' (also referred to as the intramolecular regulator of ATPase[10]) that is inserted into the cytoplasmic funnel of the SecY channel. The loop connecting the two helices is located right above the SecY pore, suggesting that it interacts with a polypeptide chain and pushes it into the SecY pore[3].

To test these ideas, we first investigated which residues of SecA's two-helix finger are important for translocation. Because mutations in this region compromise translocation[10–12], we performed a systematic analysis by changing residues 788–803 individually to alanines. These residues span the carboxy terminus of the first helix of the finger domain (residues 788–791), the entire loop between the helices (residues 792–798), and the amino terminus of the second helix (residues 799–803). The SecA mutants were purified and tested for translocation of the substrate proOmpA. *In vitro* synthesized ^{35}S-labelled proOmpA was incubated with *E. coli* SecA, ATP and proteoliposomes containing purified *E. coli* SecY complex, and translocation of proOmpA was determined by protease digestion (Fig. 1a; quantification in Fig. 1b). The results show that two residues in the helices are particularly important for translocation: Leu 791 at the end of the first helix (Fig. 1a, lane 9 versus lane 1), which probably interacts with the SecY loop between transmembrane (TM) segments 6 and 7 (ref. 3), and Pro 799 at the beginning of the second helix (lane 23), which may serve to cap the helix. Glu 802 and Tyr 803 in the second helix are also required for full activity (lanes 29 and 31), perhaps because they stabilize the structure of the two-helix finger[13]. Within the loop, the only important residue is Tyr 794 (Fig. 1a, lane 15; Fig. 1b). Similar results were obtained with cysteine mutations in the loop region (Supplementary Fig. 1). We then mutated the tyrosine residue to different amino acids. The translocation rates obtained with purified SecA mutants show that the tyrosine can be replaced by a few other amino acids with bulky, hydrophobic side chains (Tyr > Phe > Leu ≅ Trp ≅ Met > His > Val), whereas all amino acids with small or charged side chains have very low activity (Fig. 1c). The tyrosine residue is found in 80.3% of all SecA proteins (601 out of 748 SecA proteins listed in the database), but other bulky, hydrophobic residues are occasionally observed (Supplementary Tables 1 and 2; the infrequently found histidines might be uncharged at appropriate pH). Interestingly, 18 listed SecA proteins have small or hydrophilic amino acids at the critical position that would be expected to cause low translocation activity, but in all but one case (a SecA with an arginine) there is another SecA in the same organism with a conventional bulky, hydrophobic residue (Supplementary Table 3). In *Streptococcus gordonii*, the non-conventional SecA (SecA2) has a threonine and cooperates with a specialized SecY channel and other factors to export just one substrate, a 3,072-residue, serine-rich cell surface adhesin that is glycosylated in the cytosol before secretion; the conventional SecA1 has a tyrosine and is responsible for the export of all other substrates[14]. A similar situation may apply to other bacteria that contain both a conventional and non-conventional SecA[15].

Next we used disulphide crosslinking to investigate whether the loop at the tip of the two-helix finger of SecA interacts with a polypeptide chain during translocation. Because SecA interacts rather promiscuously with polypeptide chains in solution (Y.N. and K.J.E., unpublished results), we developed a strategy that allowed us to study interactions of a truly translocating polypeptide chain. We introduced two cysteines into the substrate proOmpA, one that would crosslink to the SecY channel, and another one that would crosslink to SecA (Fig. 2a). Double crosslinks of proOmpA to both SecY and SecA would indicate interactions of a polypeptide chain that has engaged both translocation components.

We first determined the positions at which cysteines in proOmpA would crosslink to a cysteine introduced into the pore ring of *E. coli*

[1]Howard Hughes Medical Institute and Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA. †Present address: Department of Pathology, Feinberg School of Medicine, Northwestern University, 303 East Chicago Avenue, Chicago, Illinois 60611, USA.

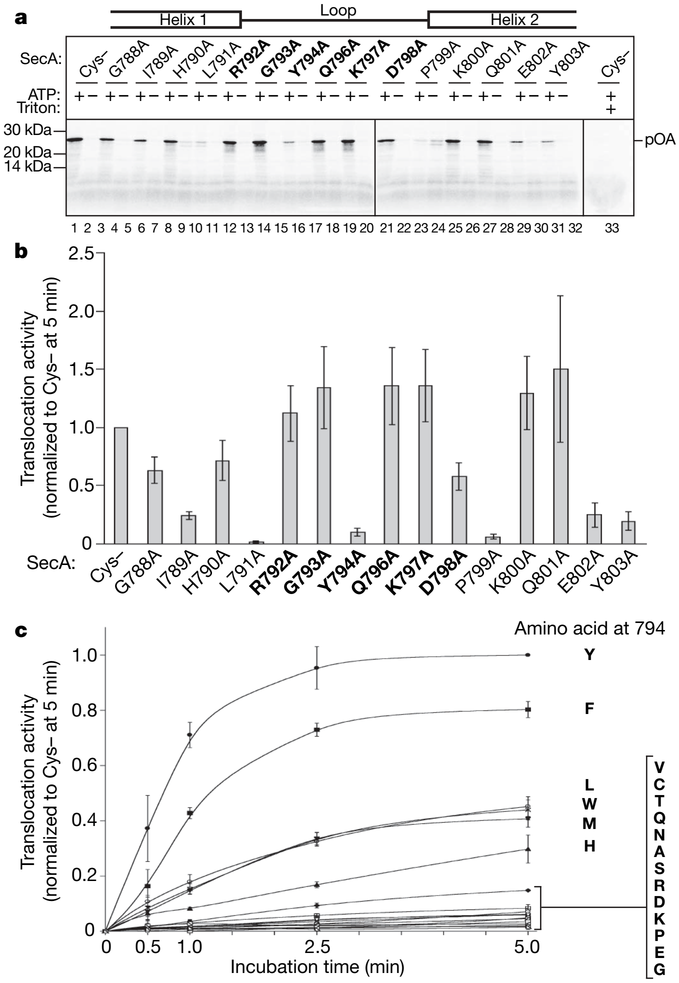


Figure 1 | An essential tyrosine at the tip of the two-helix finger. **a**, Residues in SecA's two-helix finger were individually mutated to alanines. Position 795 is an alanine in the wild type and cysteine-lacking (Cys−) proteins and therefore remained unchanged. Residues in the helices and loops are indicated in the scheme on top and loop residues are also highlighted in bold. The mutants were purified and tested for translocation activity by incubation for 5 min at 37 °C in the presence or absence of ATP with *in vitro* synthesized ^{35}S-labelled proOmpA (pOA) and proteoliposomes containing purified SecY complex. After treatment with proteinase K, the samples were separated by SDS–PAGE and analysed by autoradiography. The Cys− mutant served as a positive control. For each mutant, samples were also treated with protease in the presence of Triton X-100 to disrupt the membrane (shown here in lane 33 for the Cys− mutant). **b**, Quantification of three experiments performed as in **a** (mean and standard errors). The data were normalized with respect to the Cys− mutant. **c**, Translocation kinetics of SecA mutants in which Tyr 794 was replaced with other residues (in one-letter amino acid code). Shown are the mean and standard deviation of three experiments, normalized to the 5-min data point of the Cys− mutant, which has the wild-type tyrosine (Y) in the loop.

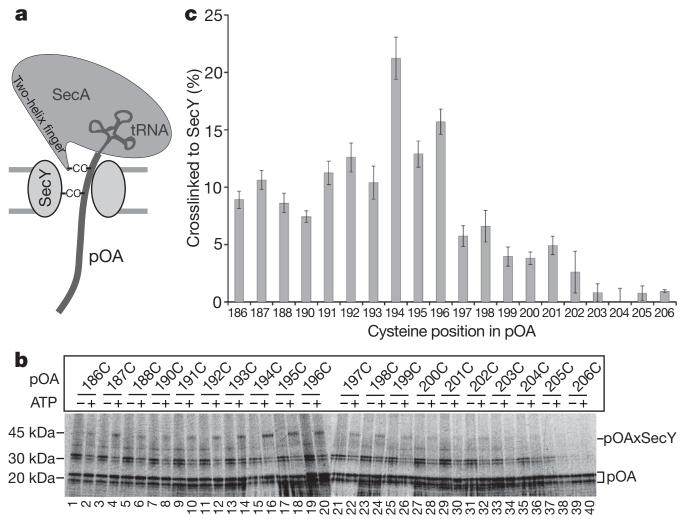


Figure 2 | Contact of a translocation intermediate with the pore of SecY. **a**, Scheme of the crosslinking strategy. A tRNA-associated fragment of ^{35}S-proOmpA (pOA) was synthesized *in vitro* and translocated by SecA into the SecY channel. The bulky tRNA prevents complete translocation. Two cysteines (C) were introduced into proOmpA, one for crosslinking to a cysteine in the two-helix finger of SecA and one for crosslinking to a cysteine in SecY. **b**, Translocation substrates (proOmpA–tRNA) of 206 residues containing single cysteines at the indicated positions were incubated in the absence or presence of ATP with a cysteine-free SecA and proteoliposomes containing purified SecY complex. SecY carried a single cysteine in the pore ring at position 282. After oxidation with Cu^{2+}-phenanthroline, the samples were treated with NEM and RNaseA, and analysed by non-reducing SDS–PAGE and autoradiography. The positions of free and crosslinked proOmpA (pOA and pOAxSecY) are indicated. **c**, Quantification of three experiments performed as in **b** (mean and standard deviation).

SecY (position 282). ProOmpA fragments of 206 amino acids, each containing a single cysteine, were synthesized by *in vitro* translation of a truncated messenger RNA in the presence of ^{35}S-methionine. With no stop codon in the mRNA, the nascent polypeptide chain remains associated with the ribosome as peptidyl-transfer RNA, and can be released from the ribosome by treatment with urea (proOmpA–tRNA)[16]. These proOmpA–tRNA substrates were then incubated with *E. coli* SecA lacking the endogenous cysteines[17,18], ATP, and proteoliposomes containing purified SecY complex. The bulky tRNA at the C terminus of the substrates prevents their complete translocation into the vesicles, thus generating translocation intermediates (Fig. 2a)[19]. An oxidant was added and the formation of a disulphide bridge was analysed by non-reducing SDS–polyacrylamide gel electrophoresis (PAGE) and autoradiography (Fig. 2b). Crosslinks between proOmpA and the pore residue of SecY were only seen in the presence of ATP, when the polypeptide chain was translocated by SecA into the channel (Fig. 2b). The strongest crosslinks were seen with cysteines at positions 186–196 of proOmpA (lanes 1–20; see quantification in Fig. 2c); residues that were closer than ten residues from the C terminus gave only weak or no crosslinks, consistent with the idea that the tRNA moiety prevents the final residues from moving into the pore.

ProOmpA residues not yet translocated into the SecY pore may be expected to interact with the two-helix finger of SecA, which is located right above the pore entrance[3] (Fig. 2a). We therefore used a proOmpA–tRNA substrate that contained a cysteine at position 194 for crosslinking to the pore position 282 in SecY (see Fig. 2b), and a cysteine at position 201 for crosslinking to position 797 near the tip of the two-helix finger of SecA. After incubation in the presence or absence of ATP, disulphide bridge formation was induced with an oxidant. A prominent double crosslink was generated (Fig. 3a, lane 8), in addition to the single crosslinks to SecY and SecA. The double crosslinks to SecY and SecA and the single crosslinks to SecY were dependent on the presence of ATP, whereas crosslinks to SecA were nucleotide independent (Fig. 3a, lane 8 versus 7), reflecting the promiscuous interaction of SecA with non-translocating substrate. Double crosslinks to SecY and SecA were not observed when either SecY or SecA lacked a cysteine (Fig. 3a, lanes 1–6). The identity of the crosslinked band was further confirmed by immunoprecipitation experiments; the high-molecular-mass band could be precipitated with either SecA or SecY antibodies (Fig. 3b, lanes 6 and 8).

To confirm that the N- and C-terminal cysteines of proOmpA contact SecY and SecA, respectively, we varied the spacing between the two cysteines in proOmpA (Fig. 3c). As before, we used a SecA mutant with a cysteine at position 797 in the two-helix finger and a SecY mutant with a cysteine at position 282. The proOmpA substrates contained a cysteine at position 194 and a second cysteine

in a more C-terminal position. The samples were incubated in the absence or presence of ATP, the membranes were sedimented, and disulphide bridge formation was induced with an oxidant. Double crosslinks were observed when the second cysteine in proOmpA was placed at positions 201–205 (Fig. 3c, lanes 7–17), indicating that the final residues of proOmpA contact SecA's two-helix finger. Positions further towards the N terminus did not give double crosslinks (positions 198–200; lanes 1–6), consistent with the expectation that they have already passed the two-helix finger of SecA and are located further inside the SecY channel.

Next, we determined the region of the two-helix finger that contacts the proOmpA substrate. We placed single cysteines at various positions in the finger of SecA and used a proOmpA substrate with cysteines at positions 194 and 201. The most prominent double crosslinks were observed with positions 795, 797 and 798 (Fig. 3d, lanes 8, 10 and 12). These positions are located at the tip of the two-helix finger, precisely where a polypeptide chain is predicted to pass according to the SecA–SecY structure[3]. Positions in the two helices, which are further away from the fingertip, did not give crosslinks (Fig. 3d, lanes 4, 14 and 16). Although some of the SecA mutants translocated proOmpA more slowly than wild-type SecA (Supplementary Fig. 1), the incubation time was sufficient to generate translocation intermediates, as indicated by the appearance of ATP-dependent SecY crosslinks (Fig. 3d). Replacement of the tyrosine at the fingertip with cysteine severely compromised translocation (Supplementary Fig. 1), and therefore only weak SecY and no SecY–SecA double crosslinks were observed (Fig. 3d, lane 6). However, because the mutant is not entirely inactive, in some experiments weak SecY and double crosslinks were seen (data not shown), indicating that this position also contacts the polypeptide chain. Taken together, these results indicate that a translocating polypeptide chain passes from the tip of the two-helix finger into the SecY pore. Because the double crosslinks could be generated when the two cysteines in proOmpA were only seven residues apart (Fig. 3c; residues 201 and 194), the translocating polypeptide chain must be in an extended conformation to span the distance of 20–25 Å between the fingertip of SecA and the pore ring of SecY. Given that even the most C-terminal residues of proOmpA can crosslink to the two-helix finger, it seems that most of the polypeptide chain has been pushed into the channel.

Our crosslinking experiments show that a translocating polypeptide chain passes directly from the tip of the two-helix finger into the SecY pore (Fig. 4). The polypeptide chain is probably positioned above the pore entrance by a clamp generated by the PPXD, the NBD2 and the HSD[3] (Fig. 4). We propose that the finger domain moves up and down during the ATP hydrolysis cycle and pushes the polypeptide into the SecY channel. The finger could interact with the polypeptide chain on ATP binding to SecA, and release the chain and reset after ATP hydrolysis. In support of this mechanism, the ADP-bound state of SecA allows a polypeptide chain to slide back into the cytosol[20]. We propose that the essential tyrosine (or other bulky, hydrophobic residue) in the loop at the tip of the finger domain contacts and drags the polypeptide chain, probably by interacting with amino acid side chains in the substrate. In the case of non-conventional SecA proteins (SecA2 proteins), a small or hydrophilic amino acid might make specific contact with special features in their unique substrates, while at the same time preventing the transport of ordinary substrates. The proposed mechanism of SecA function is analogous to the one proposed for hexameric RecA-like ATPases, such as ClpX, ClpA, HslU, the 19S subunit of the proteasome, FtsH and p97 (refs 21–25). In these cases, each monomer has a loop with a conserved tyrosine at its tip (tryptophan or phenylalanine in the case of p97 and FtsH, respectively), which contacts the polypeptide and transports it by moving up and down inside the central pore. As in the case of SecA, the aromatic residue in the loop can be

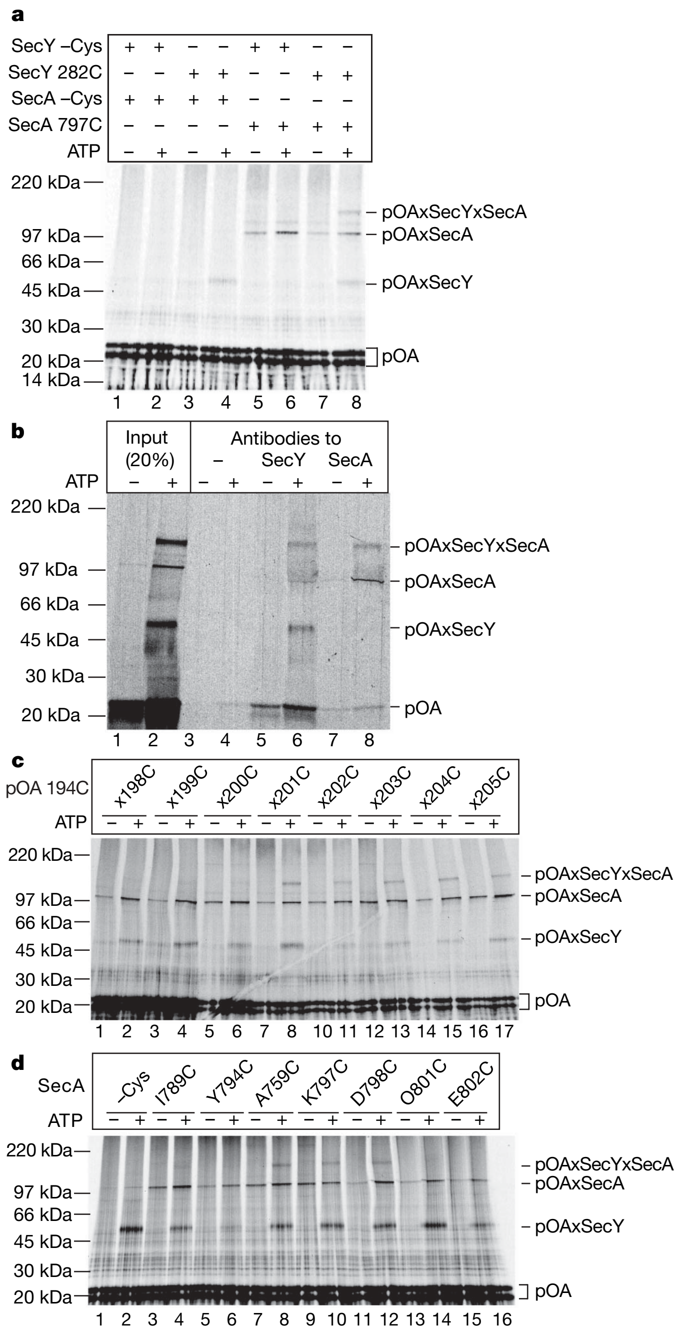


Figure 3 | The two-helix finger of SecA interacts with a translocating substrate. a, A translocation substrate (proOmpA–tRNA) of 206 residues containing cysteines at positions 194 and 201 was incubated with SecA and proteoliposomes containing SecY complex in the absence or presence of ATP. The SecA proteins either lacked a cysteine or contained a cysteine at the tip of the two-helix finger (position 797). SecY lacked a cysteine or contained a cysteine in the pore ring (position 282). After oxidation with Cu^{2+}-phenanthroline, the samples were treated with NEM and RNaseA, and analysed by non-reducing SDS–PAGE and autoradiography. The positions of free and crosslinked proOmpA (pOA, pOAxSecY, pOAxSecA and pOAxSecYxSecA) are indicated. **b**, As in **a**, but the proteoliposomes were sedimented before addition of the oxidant. SecA contained a cysteine at position 797. After solubilization in DDM, 20% of each sample was analysed directly (lanes 1 and 2), whereas the remainder was incubated with protein G beads either without antibody or with SecA or SecY antibodies (lanes 3–8). **c**, As in **a**, but the proteoliposomes containing SecY with a cysteine at position 282 were incubated with a SecA mutant containing a single cysteine at position 797 and proOmpA–tRNA with a cysteine at position 194 and a second cysteine at the indicated position. The samples were sedimented before addition of the oxidant. **d**, As in **c**, but the proteoliposomes containing SecY with a cysteine at position 282 were incubated with proOmpA–tRNA containing cysteines at positions 194 and 201 and SecA mutants containing single cysteines at the indicated positions.

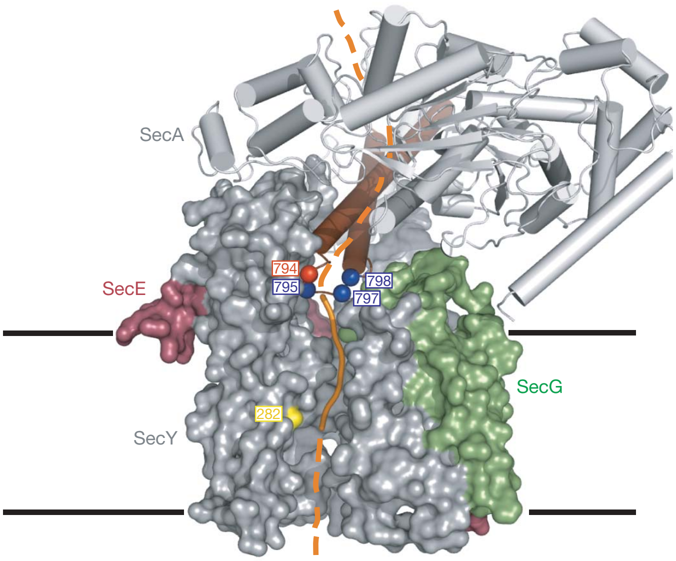


Figure 4 | Model of SecA translocating a polypeptide into the SecY channel. A segment of a translocation substrate was modelled in orange in the *Thermotoga maritima* SecA–SecY complex structure (TM2b and residues 72–76 were removed for clarity)[3]. The segment from the tip of the two-helix finger of SecA to the pore ring of SecY is shown as a solid line; the other segments as broken lines. Translocating proOmpA containing two cysteines could be double-crosslinked to position 282 of SecY (yellow) and to SecA positions in blue at the tip of the two-helix finger (brown). The essential tyrosine in the loop between the two helices is shown in red. SecY is shown as a space-filling model and SecA as a cartoon. The numbers correspond to the positions in *E. coli* SecA and SecY. The membrane boundaries are indicated.

changed to other bulky, hydrophobic residues, but not to small amino acids[22,23,26,27]. These similarities suggest that polypeptide movement occurs by a conserved, general mechanism.

METHODS SUMMARY

Point mutants of the N95 fragment of *E. coli* SecA, which lacks the non-essential C terminus, were generated by polymerase chain reaction (PCR) mutagenesis. The proteins were C-terminally His-tagged and purified by Ni^{2+}-chelating chromatography. SecY complexes were purified[19] and reconstituted with *E. coli* polar lipids into phospholipid vesicles[19,28]. Cysteine mutants were generated in a proOmpA fragment that lacked residues 175–296 and all endogenous cysteines. Fragments of 206 amino acids in length (numbering for the deletion construct) were synthesized by *in vitro* translation in rabbit reticulocyte lysate in the presence of ^{35}S-methionine. The substrates (proOmpA–tRNA) were precipitated and dissolved in 8 M urea. For crosslinking, proOmpA–tRNA was incubated for 15 min at 37 °C with SecY reconstituted into proteoliposomes, SecA and ATP[19]. In indicated cases, the membranes were sedimented before adding the oxidant. Oxidation was performed with Cu^{2+}-phenanthroline, followed by *N*-ethylmaleimide (NEM) and RNaseA treatment. After solubilization in 1% dodecylmaltoside (DDM), the samples were subjected to non-reducing SDS–PAGE. For immunoprecipitation, the solubilized samples were incubated with protein G beads containing pre-bound SecY or SecA antibodies[6]. Translocation assays were performed as described[20].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 June; accepted 19 September 2008.

1. Rapoport, T. A. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* 450, 663–669 (2007).
2. Brundage, L., Hendrick, J. P., Schiebel, E., Driessen, A. J. M. & Wickner, W. The purified *E. coli* integral membrane protein SecY/E is sufficient for reconstitution of SecA-dependent precursor protein translocation. *Cell* 62, 649–657 (1990).
3. Zimmer, J., Nam, Y. & Rapoport, T. A. Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* doi:10.1038/nature07335 (this issue).
4. Economou, A. & Wickner, W. SecA promotes preprotein translocation by undergoing ATP-driven cycles of membrane insertion and deinsertion. *Cell* 78, 835–843 (1994).
5. van den Berg, B. *et al.* X-ray structure of a protein-conducting channel. *Nature* 427, 36–44 (2004).
6. Cannon, K. S., Or, E., Clemons, W. M. Jr, Shibata, Y. & Rapoport, T. A. Disulfide bridge formation between SecY and a translocating polypeptide localizes the translocation pore to the center of SecY. *J. Cell Biol.* 169, 219–225 (2005).
7. Harris, C. R. & Silhavy, T. J. Mapping an interface of SecY (PrlA) and SecE (PrlG) by using synthetic phenotypes and *in vivo* cross-linking. *J. Bacteriol.* 181, 3438–3444 (1999).
8. Tam, P. C., Maillard, A. P., Chan, K. K. & Duong, F. Investigating the SecY plug movement at the SecYEG translocation channel. *EMBO J.* 24, 3380–3388 (2005).
9. Hunt, J. F. *et al.* Nucleotide control of interdomain interactions in the conformational reaction cycle of SecA. *Science* 297, 2018–2026 (2002).
10. Karamanou, S. *et al.* A molecular switch in SecA protein couples ATP hydrolysis to protein translocation. *Mol. Microbiol.* 34, 1133–1145 (1999).
11. Jarosik, G. P. & Oliver, D. B. Isolation and analysis of dominant secA mutations in *Escherichia coli*. *J. Bacteriol.* 173, 860–868 (1991).
12. Vrontou, E., Karamanou, S., Baud, C., Sianidis, G. & Economou, A. Global co-ordination of protein translocation by the SecA IRA1 switch. *J. Biol. Chem.* 279, 22490–22497 (2004).
13. Osborne, A. R., Clemons, W. M. Jr & Rapoport, T. A. A large conformational change of the translocation ATPase SecA. *Proc. Natl Acad. Sci. USA* 101, 10937–10942 (2004).
14. Bensing, B. A., Takamatsu, D. & Sullam, P. M. Determinants of the streptococcal surface glycoprotein GspB that facilitate export by the accessory Sec system. *Mol. Microbiol.* 58, 1468–1481 (2005).
15. Siboo, I. R., Chaffin, D. O., Rubens, C. E. & Sullam, P. M. Characterization of the accessory Sec system of *Staphylococcus aureus*. *J. Bacteriol.* 190, 6188–6196 (2008).
16. Matlack, K. E. S., Plath, K., Misselwitz, B. & Rapoport, T. A. Protein transport by purified yeast Sec complex and Kar2p without membranes. *Science* 277, 938–941 (1997).
17. Matsuyama, S., Kimura, E. & Mizushima, S. Complementation of two overlapping fragments of SecA, a protein translocation ATPase of *Escherichia coli*, allows ATP binding to its amino-terminal region. *J. Biol. Chem.* 265, 8760–8765 (1990).
18. Or, E., Navon, A. & Rapoport, T. Dissociation of the dimeric SecA ATPase during protein translocation across the bacterial membrane. *EMBO J.* 21, 4470–4479 (2002).
19. Osborne, A. R. & Rapoport, T. A. Protein translocation is mediated by oligomers of the SecY complex with one SecY copy forming the channel. *Cell* 129, 97–110 (2007).
20. Erlandson, K. J., Or, E., Osborne, A. R. & Rapoport, T. A. Analysis of polypeptide movement in the SecY channel during SecA-mediated protein translocation. *J. Biol. Chem.* 283, 15709–15715 (2008).
21. Wang, J. *et al.* Crystal structures of the HslVU peptidase-ATPase complex reveal an ATP-dependent proteolysis mechanism. *Structure* 9, 177–184 (2001).
22. Siddiqui, S. M., Sauer, R. T. & Baker, T. A. Role of the processing pore of the ClpX AAA+ ATPase in the recognition and engagement of specific protein substrates. *Genes Dev.* 18, 369–374 (2004).
23. Hinnerwisch, J., Fenton, W. A., Furtak, K. J., Farr, G. W. & Horwich, A. L. Loops in the central channel of ClpA chaperone mediate protein binding, unfolding, and translocation. *Cell* 121, 1029–1041 (2005).
24. DeLaBarre, B., Christianson, J. C., Kopito, R. R. & Brunger, A. T. Central pore residues mediate the p97/VCP activity required for ERAD. *Mol. Cell* 22, 451–462 (2006).
25. Martin, A., Baker, T. A. & Sauer, R. T. Diverse pore loops of the AAA+ ClpX machine mediate unassisted and adaptor-dependent recognition of ssrA-tagged substrates. *Mol. Cell* 29, 441–450 (2008).
26. Yamada-Inagawa, T., Okuno, T., Karata, K., Yamanaka, K. & Ogura, T. Conserved pore residues in the AAA protease FtsH are important for proteolysis and its coupling to ATP hydrolysis. *J. Biol. Chem.* 278, 50182–50187 (2003).
27. Park, E. *et al.* Role of the GYVG pore motif of HslU ATPase in protein unfolding and translocation for degradation by HslV peptidase. *J. Biol. Chem.* 280, 22892–22898 (2005).
28. Collinson, I. *et al.* Projection structure and oligomeric properties of a bacterial core protein translocase. *EMBO J.* 20, 2462–2471 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. DeLaBarre and B. Burton for critical reading of the manuscript, and R. Sauer, T. Baker and A. Horwich for discussion. The work was supported by an NIH grant. T.A.R. is a HHMI investigator. Y.N. was supported by the Damon Runyon Cancer Research Foundation (DRG 1953-07).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to T.A.R. (tom_rapoport@hms.harvard.edu).

METHODS

Mutagenesis and protein purification. Point mutants of *E. coli* SecA were generated by using the cysteine-lacking N95 fragment with a C-terminal His tag (pET30b, Novagen) as the template for QuickChange (Stratagene) site-directed mutagenesis. SecA expression in BL21 (DE3) cells was induced with 1 mM isopropyl-β-D-thiogalactoside for 3 h at 37 °C. The proteins were purified with a Ni^{2+}-chelating column, and dialysed against 50 mM Tris, pH 7.5, 100 mM NaCl, and 5 mM dithiothreitol (DTT)[13]. SecY complexes were purified as before[19] and reconstituted with *E. coli* polar lipids into phospholipid vesicles, as described[19,28].

All endogenous cysteines of proOmpA were mutated and an internal fragment encoding residues 175–296 was deleted. For crosslinking, a fragment of 206 amino acids in length (numbering for the deletion construct) was generated by PCR with a 5′ primer containing the SP6 promoter sequence and a 3′ primer lacking a stop codon. After *in vitro* transcription, proOmpA fragments were translated in rabbit reticulocyte lysate in the presence of ^{35}S-methionine (Perkin Elmer). The proteins were precipitated with three volumes of 80% saturated ammonium sulphate (pH 7.5), pelleted at 22,000*g* for 15 min, and resuspended in the original volume with 8 M urea buffered with 50 mM HEPES/KOH, pH 7.0.

Crosslinking. Crosslinking was performed in 25 µl reactions containing 50 mM HEPES/KOH, pH 7.5, 50 mM KCl, 5 mM $MgCl_2$, 0.5 mg ml^{-1} BSA and 2.5 mM ATP in the presence of 27 µg ml^{-1} of SecY proteoliposomes, 40 µg ml^{-1} of SecA and 1 µl of proOmpA–tRNA[19]. The samples were incubated for 15 min at 37 °C to form a translocation intermediate. For the indicated double-crosslinking experiments, the membranes were sedimented in an Eppendorf microcentrifuge for 15 min at 14,000 r.p.m. at 4 °C. Crosslinking was performed with 50 µM Cu^{2+}-phenanthroline for 10 min at 37 °C, and the samples were treated with NEM and RNaseA and solubilized in 1% DDM. After centrifugation for 3 min at 14,000 r.p.m. at room temperature, SDS sample buffer was added and the samples were subjected to SDS–PAGE in 4–20% Tris-HCl gels (Bio-Rad). For immunoprecipitation, the DDM-solubilized samples were mixed with 20 µl of protein-G-agarose beads pre-bound to SecY or SecA antibodies[6]. Bound material was eluted in sample buffer.

Translocation assays. A proOmpA fragment, lacking both endogenous cysteines and an internal fragment encoding residues 175–296 but containing the natural stop codon, was synthesized *in vitro*. The substrate was precipitated by ammonium sulphate and dissolved in urea. Translocation assays of 50 µl contained 50 mM Tris-HCl, pH 7.9, 50 mM KCl, 50 mM NaCl, 2.5 mM ATP, 5 mM $MgCl_2$, 1 mM DTT, 0.1 mg ml^{-1} BSA, 20 µg ml^{-1} SecA, 36 µg ml^{-1} SecY proteoliposomes and 1 µl of proOmpA. For the experiments in the absence of ATP, 22 µg ml^{-1} hexokinase and 20 mM glucose were added. After incubation for different time periods at 37 °C, the samples were incubated with 0.1 mg ml^{-1} proteinase K for 40 min on ice. The samples were treated with 2 mM phenylmethanesulphonyl fluoride and precipitated with 10% trichloroacetic acid before SDS–PAGE.

Vol 455 | 16 October 2008 | doi:10.1038/nature07421

LETTERS

Conformational transition of Sec machinery inferred from bacterial SecYE structures

Tomoya Tsukazaki[1]*, Hiroyuki Mori[2]*, Shuya Fukai[1]†, Ryuichiro Ishitani[3], Takaharu Mori[4], Naoshi Dohmae[5], Anna Perederina[6], Yuji Sugita[4], Dmitry G. Vassylyev[6], Koreaki Ito[2] & Osamu Nureki[1,3]

Over 30% of proteins are secreted across or integrated into membranes. Their newly synthesized forms contain either cleavable signal sequences or non-cleavable membrane anchor sequences, which direct them to the evolutionarily conserved Sec translocon (SecYEG in prokaryotes and Sec61, comprising α-, γ- and β-subunits, in eukaryotes). The translocon then functions as a protein-conducting channel[1]. These processes of protein localization occur either at or after translation. In bacteria, the SecA ATPase[2,3] drives post-translational translocation. The only high-resolution structure of a translocon available so far is that for SecYEβ from the archaeon *Methanococcus jannaschii*[4], which lacks SecA. Here we present the 3.2-Å-resolution crystal structure of the SecYE translocon from a SecA-containing organism, *Thermus thermophilus*. The structure, solved as a complex with an anti-SecY Fab fragment, revealed a 'pre-open' state of SecYE, in which several transmembrane helices are shifted, as compared to the previous SecYEβ structure[4], to create a hydrophobic crack open to the cytoplasm. Fab and SecA bind to a common site at the tip of the cytoplasmic domain of SecY. Molecular dynamics and disulphide mapping analyses suggest that the pre-open state might represent a SecYE conformational transition that is inducible by SecA binding. Moreover, we identified a SecA–SecYE interface that comprises SecA residues originally buried inside the protein, indicating that both the channel and the motor components of the Sec machinery undergo cooperative conformational changes on formation of the functional complex.

SecY or Sec61α, the core component of the translocon, has ten transmembrane (TM1–TM10), six cytoplasmic (C1–C6) and five periplasmic/luminal (P1–P5) regions[5] (see Supplementary Fig. 1). The amino-terminal (TM1–TM5) and carboxy-terminal (TM6–TM10) halves of SecY are assembled into a pseudo-symmetrical structure, having an hourglass-shaped channel in its interior that is gated from the periplasmic side by a 'plug' helix[4]. The translocon also has a lateral gate, through which a pre-protein's signal peptide may enter the channel to initiate the translocation process, and the hydrophobic segment of a membrane protein may exit into the lipid phase to establish a transmembrane configuration. Although the reported *M. jannaschii* SecYEβ structure is in a closed, resting state, the channel that is formed within a single heterotrimer indeed mediates translocation[6–8]. Nevertheless, the fully functional translocon might require oligomerization of the SecY (Sec61) complex[9], which provides the binding platform for the cytoplasmic partners (such as the ribosome[10,11] and SecA[12]) or forms a larger channel[13]. In the dimer model of the *Escherichia coli* SecYEG–SecA system[12], one copy of SecYEG provides a docking site for SecA, whereas another copy is used as a translocating pore. SecA[2,3,5] interacts with SecYEG[14], partially activating its ATPase[15] ('membrane ATPase'), which is then enhanced fully by a pre-protein ('translocation ATPase') during active, ongoing translocation. The C4 and C5 regions of SecY (see Supplementary Fig. 1), which are critical for the activation of SecA[5,16,17], contain residues that contact the ATPase domain of SecA[12,18]. Although SecA is believed to function through large conformational changes[19], its functional oligomeric state[9,20] (see Supplementary Discussion), as well as the molecular details of the motor function, remain elusive. We chose to study the Sec machinery from *Thermus thermophilus*[21], including the crystal structure determination of SecA[20] and SecYE, the subject of this report.

Initial crystals of *T. thermophilus* SecYE showed a resolution limit of ~6 Å. Resolution was improved when SecYE was in complex with the Fab fragment of a monoclonal antibody against *T. thermophilus* SecY. The crystal structure was determined by the multiple anomalous dispersion (MAD) method, in which SecYE had been labelled with selenomethionine (Supplementary Table 1). The refined model, at a 3.2 Å resolution, includes most of the residues of Fab, SecY and SecE (Fig. 1). The electron density map displays most of the amino acid side chains of these components (Supplementary Fig. 2). The overall architecture of SecYE is similar to that of *M. jannaschii* SecYEβ[4] (Supplementary Fig. 3), in that it has an inverted, pseudo-symmetrical arrangement of transmembrane helices, an hourglass-shaped conduit with the 'pore ring' constriction (Supplementary Fig. 2b), and a plug helix. The cytoplasmic C4–C5 loops, to which SecA and the ribosome bind[4,10–12,18], protrude prominently from the membrane-embedded region. The Fab binds tightly to the highly conserved Ile 347–Phe 359 segment in the C5 loop of SecY (Supplementary Fig. 4a).

Despite the overall similarity, a closer comparison with *M. jannaschii* SecYEβ revealed that their C-terminal halves occupy different conformations (Supplementary Figs 3 and 5). The Cα root mean squared deviation was 1.36 Å for TM1–TM5 and 2.85 Å for TM6–TM10. TM6, TM8 and TM9 of *T. thermophilus* SecYE show displacements of 5 Å, 4 Å and 6 Å, respectively, from those in *M. jannaschii* SecYEβ, expanding the TM2–TM8 distance in the lateral gate area of the former (arrows in Supplementary Fig. 5b). TM6 and TM8 are tilted by ~25° and ~10°, respectively, and TM9 shows a translational movement to the periphery of the channel (Supplementary Fig. 5c). Similar displacement of TM8 and TM9 was noted, when the SecYE structure was compared with the models of the *E. coli* SecY structure[10,22] (Supplementary Fig. 6). The altered transmembrane arrangement creates a ~6 × 15 Å hydrophobic crack formed by the cytoplasmic

[1]Department of Biological Information, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama-shi, Kanagawa 226-8501, Japan. [2]Institute for Virus Research, Kyoto University, Kyoto 606-8507, Japan. [3]Division of Structural Biology, Department of Basic Medical Science, The Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. [4]Advanced Science Institute, [5]Biomolecular Characterization Team and CREST/JST, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan. [6]Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham, Schools of Medicine and Dentistry, 402B Kaul Genetics Building, 720 20th Street South, Birmingham, Alabama 35294, USA. †Present address: Structural Biology Laboratory, Life Science Division, Synchrotron Radiation Research Organization, The University of Tokyo, 211 General Research Building, Institute of Molecular and Cellular Biosciences, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan.
*These authors contributed equally to this work.

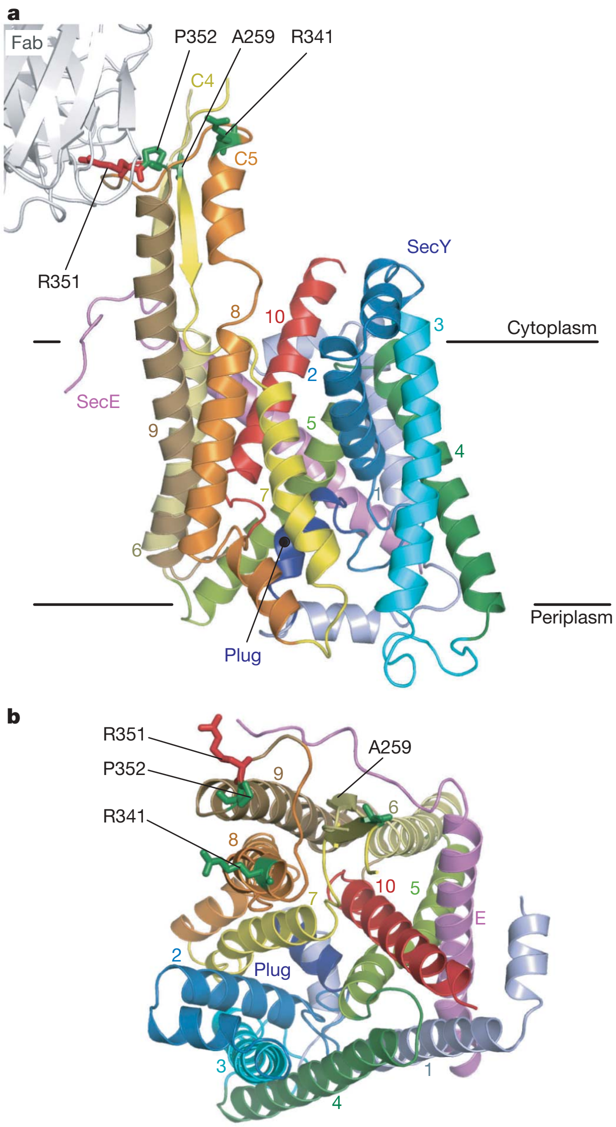


Figure 1 | Overall structure of *T. thermophilus* SecYE. a, b, The SecYE complex viewed from the lateral gate side (**a**) and the cytoplasm (**b**). The SecY transmembranes are coloured light blue to red from the N to C termini, and SecE is coloured pink. Arg 351 (ref. 17) is coloured red and is shown in stick representation. The residues coloured green in stick representation were mutated to cysteine for intermolecular crosslinking experiments.

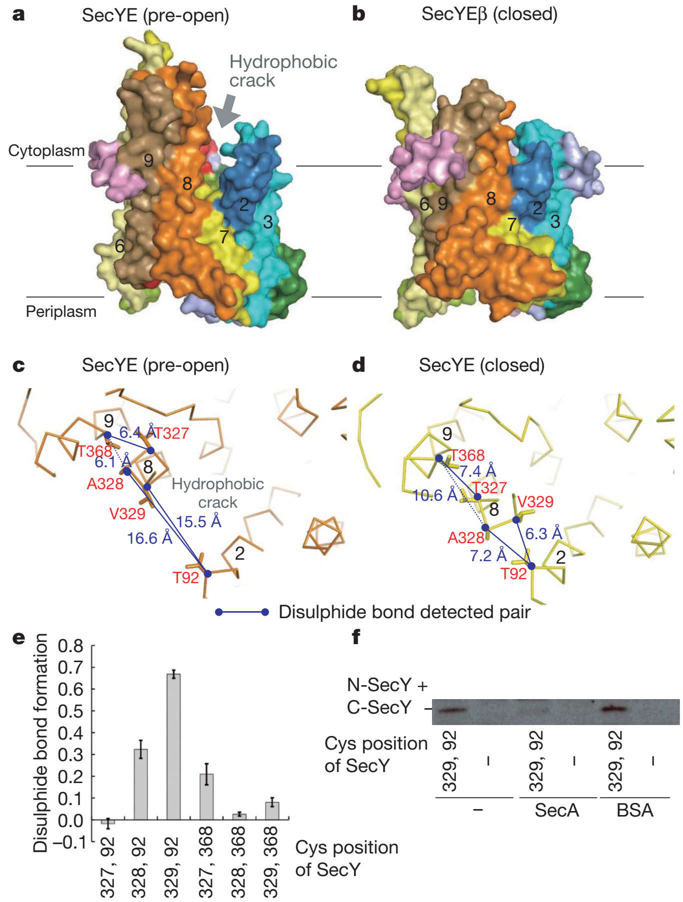


Figure 2 | Comparison of the *T. thermophilus* SecYE and *M. jannaschii* SecYEβ structures. a, b, Molecular surfaces of SecYE (**a**) and SecYEβ (**b**), coloured as in Fig. 1a. Transmembrane regions are numbered. **c, d,** The cytoplasmic regions of TM2, TM8 and TM9 of SecY. Pre-open (**c**, crystal structure of Fab–SecYE) and closed forms (**d**, without Fab, molecular dynamics analysis at 72.93 ns) are shown. Numbers show distances between α carbons. **e,** Intramolecular disulphide bond formation in SecY as assessed by quantitative carboxymethylation. Averages of three analyses are shown with s.d. **f,** SecA-mediated inhibition of SecY (Thr92Cys–Val329Cys) intramolecular disulphide bond formation in the presence of AMP-PNP.

segments of TM2, TM7 and TM8 (Fig. 2a and Supplementary Fig. 7). SecYE with an exposed crack could represent an alternative conformation of the translocon that we term 'pre-open' as opposed to or to distinguish it form the closed state of *M. jannaschii* SecYEβ.

To address whether *T. thermophilus* SecY undergoes conformational transition between the pre-open and closed states, we performed a molecular dynamics simulation (Supplementary Discussion). During the 100-ns simulation, SecYE underwent large conformational changes, in which TM8 and TM9 moved from that of the open state towards the closed state (Supplementary Fig. 8 and Supplementary Movies 1 and 2). The low (6.0 Å) resolution crystal structure of the antibody-free SecYE (Supplementary Table 1 and Supplementary Discussion) also showed a tendency towards the closed form, with the swinging and bending of TM8 (Supplementary Fig. 9). Taken together, the closed form seems to represent the energetically favoured, ground state of SecYE in the absence of any interacting components. In turn, it is conceivable that the pre-open form is a conformational substate induced by the antibody binding and possibly by the binding of a physiological ligand, SecA.

T. thermophilus SecA binds to the same (Pro-Gly-Ile-Arg-Pro-Gly) motif in C5 (see Fig. 3a) as Fab (Supplementary Fig. 4a), which inhibited the *in vitro* translocation activity of the translocon (Supplementary Fig. 4b). The arrangements of the transmembrane segments of SecYE were studied by intramolecular disulphide bond formation between cysteines placed at six different pairs of positions (Fig. 2c, d), which was assessed by reductive and non-reductive carboxymethylation. These data showed that SecY(Thr92Cys–Ala328Cys), SecY(Thr92Cys–Thr329Cys) and SecY(Thr327Cys–Val368Cys) form intramolecular disulphide bonds (Fig. 2e). The 92–328 and 92–329 linkages suggest that the isolated SecYE predominantly assumes the closed conformation, rather than the pre-open form (Fig. 2c, d). SecA decreased the 92–329 disulphide bond formation in the membrane-embedded SecYE (Fig. 2f and Supplementary Fig. 10). This would be consistent with a notion that SecA facilitates the conformational transition of SecYE from the closed to the pre-open states. The swinging of TM8 might initiate the opening of the channel gate through exposure of the hydrophobic crack (Fig. 2a, b and Supplementary Fig. 3). The crack contains some evolutionarily conserved hydrophobic residues (Ile 85, Pro 273, Phe 276, Ala 277, Phe 322 and Tyr 326) (Supplementary Figs 2a, 7 and 11a), which our mutation studies have shown to be functionally important (Supplementary Discussion and Supplementary Fig. 11b, c). Considering the observation that a signal peptide contacts TM2, TM7

and TM8 (ref. 23), it is tempting to speculate that the crack participates in the reception of signal peptides from SecA.

SecA contains two nucleotide-binding folds (NBFs; also called NBD), a pre-protein crosslinking domain (PPXD) and C-terminal translocation domains (HWD and IRA1), which are all connected by a long α-helical scaffold domain (HSD) (Fig. 3a top panel and Supplementary Fig. 12). To identify the *T. thermophilus* SecA residues that contact the C4–C5 cytoplasmic domains of SecY (refs 12, 18), we introduced single cysteines into selected positions of SecA (Supplementary Fig. 12). We first mutated residues on the NBF1 surface, as NBF1 has been shown to be in the physical proximity of SecY (ref. 12). Single cysteines were also introduced into SecY positions 259, 341 and 352 (Figs 1 and 3a) for examination of the intermolecular disulphide bond formation (Fig. 3c, Supplementary Fig. 13 and Supplementary Discussion). Two combinations, SecA(Pro202Cys)–SecY(Ala259Cys) and SecA(Leu775Cys)–SecY(Pro352Cys), formed a disulphide bond on oxidation (Fig. 3a, c, lanes 1 and 7), in which SecYE must be proteoliposome-integrated (Supplementary Fig. 13a, c, lanes 9 and 15). Thus, residues 775 (Fig. 3a, purple) and 202 (Fig. 3a, green) of SecA are adjacent to the C5 and C4 residues of SecY, respectively (see Supplementary Discussion, Supplementary Fig. 13 and ref. 12 for the specificity of molecular interactions involving SecA).

Assuming that the molecular contacts between SecA at 202 and SecY at 259 and between SecA at 775 and SecY at 352 take place simultaneously, the two cytoplasmic protrusions of SecY must be accommodated between the NBF1 and C-terminal translocation domain of SecA (Fig. 3a). However, the SecA protomer structure lacks a large enough opening in this region (Fig. 3a, arrow), raising the possibility that the SecY-associated form of SecA has undergone a conformational change. We found that an evolutionarily conserved (Supplementary Fig. 14) region ($_{180}$Gly-Phe-Asp-Tyr-Leu-Arg-Glu-Gln-Met$_{188}$) corresponding to the C-terminal half of motif IV[24] (Fig. 3b) interacts with SecY. Although the cysteines introduced into the motif IV positions of native SecA were not readily accessible to alkylation by a hydrophobic modifier (Supplementary Fig. 15 and Supplementary Discussion), they formed either a disulphide bond or a bis-maleimidoethane (BMOE; arm length of 8 Å)-mediated crosslinkage with a specific cysteine at C5 of SecY (Fig. 3d, e and Supplementary Fig. 16). Thus, the interaction with the translocon leads to the exposure of residues 182, 185, 186 and 188 on the molecular surface (Supplementary Discussion). The motif IV region indeed undergoes a conformational change that is coordinated with the formation of a motor-translocon complex.

The ATPase of SecA is tightly downregulated in the resting state (intrinsic ATPase) through the interaction between the ATPase and the IRA1 domains[25] (Fig. 3a). Our observation that binding to the translocon physically separates NBF1 and IRA1 excellently explains the translocon-mediated triggering of the membrane ATPase. The motif IV segment communicates with an anti-parallel β-sheet that is involved in the propagation of a pre-protein-binding signal to the ATPase domain[26] (Fig. 3a, b). The surface exposure and the translocon interaction of motif IV might modulate the SecA function, not only by keeping IRA1 away, but also by enabling the β-sheet to activate the ATPase further, in response to pre-protein binding (translocation ATPase).

SecA interacts with SecY in at least two different modes[18]: the one involving the SecY C4–C5 domains is probably required for the ATPase activation, whereas the other, involving C6, accompanies the actual SecA-driven translocation. The SecA–SecY interface we identified in this study seems to correspond to the first interaction

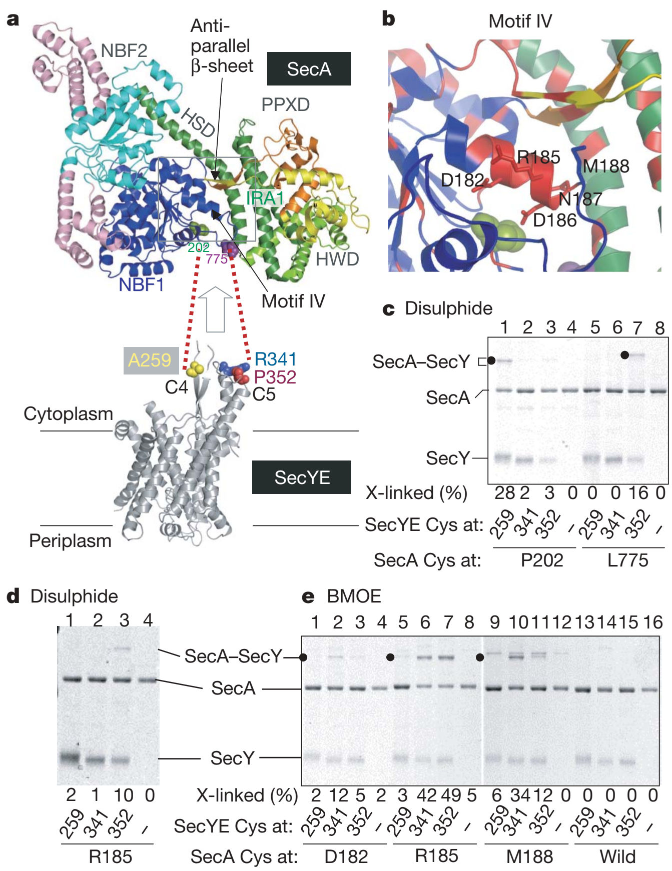


Figure 3 | Contacting residues between *T. thermophilus* SecA and SecYE. **a**, The SecA structure (Protein Data Bank 2IPC) is colour-coded for its domains[30], except that light pink indicates *T. thermophilus* SecA-specific regions. SecY contact residues identified in this study are space filled in green and purple. **b**, Close-up view of the α-helix in motif IV; evolutionarily conserved regions are shown in red, and the positions replaced by cysteines are indicated in stick representation. **c**, Disulphide crosslinking between SecA and SecY. The disulphide complex is indicated by solid circles. **d**, **e**, Disulphide and BMOE crosslinking of SecA and SecY.

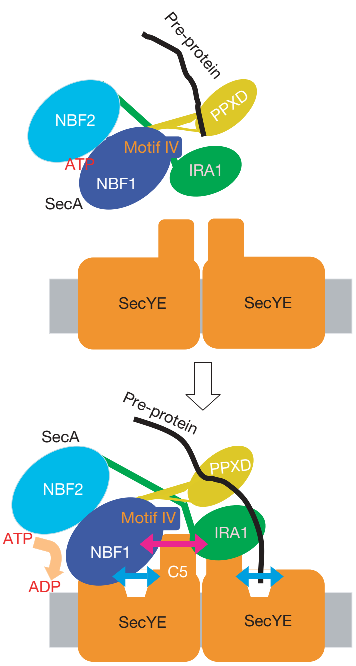


Figure 4 | Multiple modes of SecA–SecY interactions. According to the dimer model[12], one copy of SecY serves as a SecA-docking site, and the other functions as a translocation pore. The SecA–SecY interaction observed here should represent the one between the non-translocating copy of SecY and SecA, and is crucial for the SecA ATPase activation. Both the SecA and SecY components undergo conformational changes on their interaction, as shown by bidirectional arrows. The orientation of SecY protomers in the dimeric assembly is shown arbitrarily.

mode. According to the SecY dimer model by ref. 12, the non-translocating copy of SecY requires the essential Arg 357 residue[17]. The SecA–SecY interaction that we observed here directly involves Arg 351 (counterpart of *E. coli* SecY Arg 357) and might correspond to the one involving the non-translocating SecY copy (Fig. 4). In view of the occurrence of the second mode of SecA binding to the translocon, the translocating SecY copy should also interact with SecA.

The binding of the Fab fragment or SecA to the cytoplasmic region of SecY seems to induce the pre-open state of the translocon. The opening may occur in both copies of SecY, and have a role in the initial reception of the signal-peptide–SecA complex, which could be followed by further propagation of conformational changes within the translocon (Fig. 4). The pre-opening could also be important for the activation of the SecA ATPase (Fig. 4). Our proposal that the translocon and the translocation-driving motor undergo cooperative conformational changes should be substantiated by the structural determination of the SecA–translocon complex with a bound pre-protein substrate.

METHODS SUMMARY

The *T. thermophilus* SecYE complex was overproduced in *E. coli*, solubilized from the membrane with *n*-dodecyl-β-D-maltoside and purified by three successive chromatography steps. It was mixed with the Fab fragment of a monoclonal antibody against SecY, and the Fab–SecYE complex was isolated by gel filtration chromatography. Crystals of the complex were grown by vapour diffusion. The structure of the SecYE complex was determined by the MAD method, using the selenomethionine-labelled SecYE crystals, with refinement to R_{work}/R_{free} of 24.4%/28.0% at 3.2 Å resolution (Supplementary Table 1). Molecular graphics were generated with PyMol[27]. An all-atom molecular dynamics simulation of the pre-open form of SecYE with explicit solvent and phospholipids was performed using the NAMD2 program[28], as described in Supplementary Information.

To quantify the intramolecular disulphide bond formation in the purified double-cysteine mutants of SecYE, the mutants were subjected to reductive or non-reductive carboxymethylation in the presence of 6 M guanidine hydrochloride, as described previously[29]. For intra- or intermolecular disulphide crosslinking, the samples were treated with an oxidant, $Cu(phen)_3$, and were analysed by non-reducing SDS–PAGE. To assess the effect of *T. thermophilus* SecA on the conformation of SecYE, inverted membrane vesicles, prepared from *E. coli* cells overproducing the SecYE double cysteine mutants, were pre-incubated with SecA or BSA, as specified, before oxidation. For intermolecular crosslinking between SecYE and SecA, a single cysteine mutant of SecA was mixed with proteoliposomes containing SecYE with a single cysteine at a specific SecY position. For the BMOE-mediated crosslinking of SecY and SecA, single cysteine samples were incubated with BMOE.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 July; accepted 10 September 2008.

1. Rapoport, T. A. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* **450**, 663–669 (2007).
2. Papanikou, E., Karamanou, S. & Economou, A. Bacterial protein secretion through the translocase nanomachine. *Nature Rev. Microbiol.* **5**, 839–851 (2007).
3. Vrontou, E. & Economou, A. Structure and function of SecA, the preprotein translocase nanomotor. *Biochim. Biophys. Acta* **1694**, 67–80 (2004).
4. Van den Berg, B. *et al.* X-ray structure of a protein-conducting channel. *Nature* **427**, 36–44 (2004).
5. Mori, H. & Ito, K. The Sec protein-translocation pathway. *Trends Microbiol.* **9**, 494–500 (2001).
6. Cannon, K. S., Or, E., Clemons, W. M. Jr, Shibata, Y. & Rapoport, T. A. Disulfide bridge formation between SecY and a translocating polypeptide localizes the translocation pore to the center of SecY. *J. Cell Biol.* **169**, 219–225 (2005).
7. Tam, P. C., Maillard, A. P., Chan, K. K. & Duong, F. Investigating the SecY plug movement at the SecYEG translocation channel. *EMBO J.* **24**, 3380–3388 (2005).
8. Li, W. *et al.* The plug domain of the SecY protein stabilizes the closed state of the translocation channel and maintains a membrane seal. *Mol. Cell* **26**, 511–521 (2007).
9. Rusch, S. L. & Kendall, D. A. Oligomeric states of the SecA and SecYEG core components of the bacterial Sec translocon. *Biochim. Biophys. Acta* **1768**, 5–12 (2007).
10. Mitra, K. *et al.* Structure of the *E. coli* protein-conducting channel bound to a translating ribosome. *Nature* **438**, 318–324 (2005).
11. Menetret, J. F. *et al.* Ribosome binding of a single copy of the SecY complex: implications for protein translocation. *Mol. Cell* **28**, 1083–1092 (2007).
12. Osborne, A. R. & Rapoport, T. A. Protein translocation is mediated by oligomers of the SecY complex with one SecY copy forming the channel. *Cell* **129**, 97–110 (2007).
13. Kida, Y., Morimoto, F. & Sakaguchi, M. Two translocating hydrophilic segments of a nascent chain span the ER membrane during multispanning protein topogenesis. *J. Cell Biol.* **179**, 1441–1452 (2007).
14. Douville, K., Price, A., Eichler, J., Economou, A. & Wickner, W. SecYEG and SecA are the stoichiometric components of preprotein translocase. *J. Biol. Chem.* **270**, 20106–20111 (1995).
15. Lill, R. *et al.* SecA protein hydrolyzes ATP and is an essential component of the protein translocation ATPase of *Escherichia coli*. *EMBO J.* **8**, 961–966 (1989).
16. Mori, H. & Ito, K. Biochemical characterization of a mutationally altered protein translocase: proton motive force stimulation of the initiation phase of translocation. *J. Bacteriol.* **185**, 405–412 (2003).
17. Mori, H. & Ito, K. An essential amino acid residue in the protein translocation channel revealed by targeted random mutagenesis of SecY. *Proc. Natl Acad. Sci. USA* **98**, 5128–5133 (2001).
18. Mori, H. & Ito, K. Different modes of SecY-SecA interactions revealed by site-directed *in vivo* photo-cross-linking. *Proc. Natl Acad. Sci. USA* **103**, 16159–16164 (2006).
19. Economou, A. & Wickner, W. SecA promotes preprotein translocation by undergoing ATP-driven cycles of membrane insertion and deinsertion. *Cell* **78**, 835–843 (1994).
20. Vassylyev, D. G. *et al.* Crystal structure of the translocation ATPase SecA from *Thermus thermophilus* reveals a parallel, head-to-head dimer. *J. Mol. Biol.* **364**, 248–258 (2006).
21. Mori, H. *et al.* Fluorescence resonance energy transfer analysis of protein translocase. SecYE from *Thermus thermophilus* HB8 forms a constitutive oligomer in membranes. *J. Biol. Chem.* **278**, 14257–14264 (2003).
22. Breyton, C., Haase, W., Rapoport, T. A., Kuhlbrandt, W. & Collinson, I. Three-dimensional structure of the bacterial protein-translocation complex SecYEG. *Nature* **418**, 662–665 (2002).
23. Plath, K., Mothes, W., Wilkinson, B. M., Stirling, C. J. & Rapoport, T. A. Signal sequence recognition in posttranslational protein transport across the yeast ER membrane. *Cell* **94**, 795–807 (1998).
24. Sianidis, G. *et al.* Cross-talk between catalytic and regulatory elements in a DEAD motor domain is essential for SecA function. *EMBO J.* **20**, 961–970 (2001).
25. Karamanou, S. *et al.* A molecular switch in SecA protein couples ATP hydrolysis to protein translocation. *Mol. Microbiol.* **34**, 1133–1145 (1999).
26. Karamanou, S. *et al.* Preprotein-controlled catalysis in the helicase motor of SecA. *EMBO J.* **26**, 2904–2914 (2007).
27. Delano, W. L. The PyMOL molecular graphics system. v.0. 97 <http://pymol.sourceforge.net/> (2002).
28. Kalé, L. *et al.* NAMD2: Greater scalability for parallel molecular dynamics. *J. Comput. Phys.* **151**, 283–312 (1999).
29. Shindo, N. *et al.* Separation of 18 6-aminoquinolyl-carbamyl-amino acids by ion-pair chromatography. *Anal. Biochem.* **249**, 79–82 (1997).
30. Hunt, J. F. *et al.* Nucleotide control of interdomain interactions in the conformational reaction cycle of SecA. *Science* **297**, 2018–2026 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Inaba, Y. Akiyama and M. Hattori for useful suggestions about sample preparation and crystallization; T. Sakamoto and T. Saika for their assistance in the purification of *T. thermophilus* SecYE; K. Mochizuki, M. Sano, K. Yoshikaie, T. Adachi and Y. Echizen for technical support; and the beamline staff members at BL41XU of SPring-8 (Sayo, Japan) and NW12 of PF (Tsukuba, Japan) for technical help during data collection. We also thank I. Artsimovitch for critically reading the manuscript. This work was supported by a SORST program grant from JST (Japan Science and Technology) to O.N., by a CREST grant from JST to K.I. and N.D., by a BIRD grant from JST to H.M. and Y.S., by Global COE Program (Center of Education and Research for Advanced Genome-Based Medicine) and a grant for the National Project on Protein Structural and Functional Analyses from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) to O.N., by NIH grants to D.G.V., by grants from MEXT to H.M., S.F., R.I., K.I. and O.N., and by Mitsubishi Foundation grants to O.N.

Author Contributions T.T. carried out the structural determination and the biochemical experiments of *T. thermophilus* SecYE. H.M. carried out biochemical analyses of SecA–SecY interactions. A.P. and D.G.V. assisted with the crystallization and data collection of SecYE as well as with manuscript preparation. S.F., R.I. and O.N. assisted with the structural determination. T.M. and Y.S. performed the molecular dynamics simulation. N.D. performed disulphide-bond quantification and mass spectrometry. All authors discussed the results and commented on the manuscript. O.N. and K.I. supervised the work and wrote/edited the manuscript.

Author Information The coordinates and structure factors have been deposited in the Protein Data Bank, under the accession codes 2ZJS for Fab–SecYE and 2ZQP for SecYE. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to O.N. (nureki@ims.u-tokyo.ac.jp) or K.I. (Koreaki.Ito@ky5.ecs.kyoto-u.ac.jp).

METHODS

Purification and crystallization of SecYE. The full-length *T. thermophilus* SecY(Leu2Val/Arg252Glu) with a His_6-tag at the C terminus, as well as the full-length *T. thermophilus* SecE, were overexpressed in *E. coli*, which carried two plasmids: SecYE-overproducing pTT159 and pSTD343 (ref. 21). The SecYE complex was solubilized from the membrane fraction with *n*-dodecyl-β-D-maltoside, and was purified by three successive chromatography steps. The His_6-tag of SecY was spontaneously cleaved after storage at 20 °C for 2 weeks, and the protein was re-chromatographed on a Ni-NTA column. A hybridoma producing the monoclonal antibody against SecYE was made and identified by MBL Co. and our in-house screenings. The Fab fragment was purified by IBL Co. After mixing SecYE with a stoichiometric excess of Fab, the Fab–SecYE complex was purified by gel filtration. Crystals of the Fab–SecYE complex were grown by the vapour diffusion method. Details of the sample preparations will be described elsewhere (T.T. *et al.*, manuscript in preparation).

Crystallographic analysis. Diffraction data were collected at 100 K from beamline BL41XU at SPring-8 and from beamline NW12 at PF. The diffraction data were processed using DENZO/SCALEPACK[31] and the CCP4 suite[32]. The crystal structure of SecYE was determined by the MAD method, using the diffraction data from the selenomethionine (SeMet)-labelled SecYE complex. The heavy atom sites were identified with the program SnB[33]. The phases were calculated using the program SHARP[34]. The initial phases were improved by density modification, using the programs SOLOMON[35] and DM[36]. An initial atomic model of SecYE and Fab was manually built with the program O[37], using the structures of a Fab (PDB 2A6D) and *M. jannaschii* SecYEβ (PDB 1RH5) as references, and was then refined using CNS[38] and Refmac[32]. By iterative model building and map calculation, the structure of the Fab–SecYE(SeMet) complex was refined to R_{work}/R_{free} of 31.8%/34.9% at 3.3 Å resolution. The crystal structure of the native Fab–SecYE complex was determined by molecular replacement, using the program MolRep[39], and was refined with R_{work}/R_{free} = 24.4%/28.0% at 3.2 Å resolution (Supplementary Table 1). The quality of the model was substantially improved, using the zonal scaling[40]. The asymmetric unit contains a single Fab–SecYE complex. The refined model includes almost all of the residues of Fab, residues 1–247 and 255–422 of SecY (Supplementary Fig. 1) and residues 12–57 of SecE (60 amino acid residues in total). A Ramachandran plot with PROCHECK[41] showed that 91.8%, 7.5% and 0.8% of the residues in the SecYE model were in the most favoured, allowed and generously allowed regions, respectively. Molecular graphics were created with PyMol[27].

Preparation of SecA and SecYE variants. The SecA(Cys648Ser) mutation was first introduced into the SecA-overproducing plasmid, pHM451 (ref. 20), using QuikChange mutagenesis, to yield pHM512, which was in turn used as a template for the introduction of single cysteine substitutions. Native SecYE lacks cysteine. Similarly, single and double cysteine substitutions were introduced into selected positions of SecY, using the SecY–SecE overproducing plasmids pTT008 (ref. 21) and pTT159, respectively, as templates. SecA and SecYE variants were purified as described previously[21,42] and were functional in *in vitro* translocation assays, except for SecA(Arg185Cys) (data not shown).

Molecular dynamics simulation. An all-atom molecular dynamics simulation was carried out using the NAMD2 program[28], as described in Supplementary Information.

Quantification of intramolecular disulphide formation by SecY double-cysteine mutants. *T. thermophilus* SecY double-cysteine variants were subjected to reductive or non-reductive carboxymethylation. For reductive carboxymethylation, the protein (~10 μg) was dissolved in 100 μl of 1% dithiothreitol in 6 M guanidine hydrochloride, 1 M Tris-HCl (pH 8.5) and 10 mM EDTA, and was heated at 80 °C for 30 min. After cooling, alkylation was performed by the addition of 10 μl of a 25% iodoacetic acid solution in 1 N NaOH and an incubation at room temperature for 30 min in the dark. For non-reductive carboxymethylation, the protein was dissolved in 2.5% iodoacetic acid in 6 M guanidine hydrochloride, 1 M Tris-HCl (pH 8.5) and 10 mM EDTA. The reaction mixtures were desalted with a Sephadex G-25 syringe (1 ml), and pooled fractions of the carboxymethylated protein were dried and hydrolysed in 6 N HCl vapour at 110 °C for 20 h. The acid hydrolysate was derivatized with 6-aminoquinolyl-*N*-hydroxysuccinimidyl carbamate, and was quantified as described previously[29]. Molecular mass determinations of the prepared SecY double-cysteine variants excluded the possibility that small compounds, such as glutathione, had been attached to the sulphydryl groups (data not shown).

Intramolecular disulphide crosslinking of SecY. Inverted membrane vesicles (IMVs) were prepared from *T. thermophilus* SecYE-overproducing *E. coli* cells[21], and were suspended in buffer A (20 mM HEPES-Na (pH 7.5), 100 mM NaCl, 5% (v/v) glycerol, 5 mM $MgSO_4$ and 1 mM AMP-PNP). The IMVs (approximately 3 μg protein) were incubated with 2 mM $Cu(phen)_3$ in 100 μl of buffer A[43]. To remove the oxidant, the IMVs were collected by ultracentrifugation and resuspended in buffer A twice. Subsequently, trichloroacetic acid (final concentration of 5%) was added to precipitate proteins, which were then collected by centrifugation, washed with acetone and dissolved in 2% SDS, 50 mM Tris-HCl (pH 8), 5 mM EDTA, 10% glycerol and 5 mM iodoacetamide, by agitation at room temperature for 30 min. After SDS–PAGE under non-reducing conditions, the C-terminal portion of SecY was detected by immunoblotting using anti-SecY(C5); the original antibody we used for preparation of the Fab fragment.

Crosslinking between SecA and SecY. The reaction mixture (40 μl) contained 4 μl of proteoliposomes[21], with a *T. thermophilus* SecY(single cysteine)E variant, a purified single-cysteine SecA variant (10 ng μl^{-1}), 50 mM Tris-HCl (pH 7.5), 5 mM $MgSO_4$ and 100 mM NaCl. It was then incubated with 50 μM $Cu(phen)_3$ at 37 °C for 60 min. After quenching the oxidant with 10 mM neocuproine, the proteins were precipitated with 5% trichloroacetic acid, washed with acetone and dissolved in 50 mM Tris-HCl (pH 6.8) containing 2% SDS and 5 mM iodoacetamide, by incubation at 37 °C for 40 min with mild mixing. For BMOE crosslinking between SecA and SecY, the reaction mixture as described above was incubated with 50 μM BMOE (bis-maleimidoethane, Pierce), instead of $Cu(phen)_3$, at 37 °C for 30 min. The reaction was terminated with 25 mM dithiothreitol. The crosslinked products were separated by SDS–PAGE.

31. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
32. Collaborative Computational Project, Number 4. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
33. Weeks, C. M. & Miller, R. The design and implementation of SnB v2.0. *J. Appl. Crystallogr.* **32,** 120–124 (1999).
34. de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement in the MIR and MAD methods. *Methods Enzymol.* **276,** 472–494 (1997).
35. Abrahams, J. P. & Leslie, A. G. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52,** 30–42 (1996).
36. Cowtan, K. D. & Main, P. Phase combination and cross validation in iterated density-modification calculations. *Acta Crystallogr. D* **52,** 43–48 (1996).
37. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
38. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
39. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. in. *J. Appl. Crystallogr.* **30,** 1022–1025 (1997).
40. Vassylyev, D. G. *et al.* Structural basis for substrate loading in bacterial RNA polymerase. *Nature* **448,** 163–168 (2007).
41. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26,** 283–291 (1993).
42. Vassylyeva, M. N. *et al.* Cloning, expression, purification, crystallization and initial crystallographic analysis of the preprotein translocation ATPase SecA from Thermus thermophilus. *Acta Crystallogr. F* **62,** 909–912 (2006).
43. Veenendaal, A. K., van der Does, C. & Driessen, A. J. Mapping the sites of interaction between SecY and SecE by cysteine scanning mutagenesis. *J. Biol. Chem.* **276,** 32559–32566 (2001).

LETTERS

The type IV mucolipidosis-associated protein TRPML1 is an endolysosomal iron release channel

Xian-Ping Dong[1], Xiping Cheng[1], Eric Mills[1], Markus Delling[2], Fudi Wang[3], Tino Kurz[4] & Haoxing Xu[1]

TRPML1 (mucolipin 1, also known as MCOLN1) is predicted to be an intracellular late endosomal and lysosomal ion channel protein that belongs to the mucolipin subfamily of transient receptor potential (TRP) proteins[1–3]. Mutations in the human *TRPML1* gene cause mucolipidosis type IV disease (ML4)[4,5]. ML4 patients have motor impairment, mental retardation, retinal degeneration and iron-deficiency anaemia. Because aberrant iron metabolism may cause neural and retinal degeneration[6,7], it may be a primary cause of ML4 phenotypes. In most mammalian cells, release of iron from endosomes and lysosomes after iron uptake by endocytosis of Fe^{3+}-bound transferrin receptors[6], or after lysosomal degradation of ferritin–iron complexes and autophagic ingestion of iron-containing macromolecules[6,8], is the chief source of cellular iron. The divalent metal transporter protein DMT1 (also known as SLC11A2) is the only endosomal Fe^{2+} transporter known at present and it is highly expressed in erythroid precursors[6,9]. Genetic studies, however, suggest the existence of a DMT1-independent endosomal and lysosomal Fe^{2+} transport protein[9]. By measuring radiolabelled iron uptake, by monitoring the levels of cytosolic and intralysosomal iron and by directly patch-clamping the late endosomal and lysosomal membrane, here we show that TRPML1 functions as a Fe^{2+} permeable channel in late endosomes and lysosomes. ML4 mutations are shown to impair the ability of TRPML1 to permeate Fe^{2+} at varying degrees, which correlate well with the disease severity. A comparison of *TRPML1*$^{-/-}$ ML4 and control human skin fibroblasts showed a reduction in cytosolic Fe^{2+} levels, an increase in intralysosomal Fe^{2+} levels and an accumulation of lipofuscin-like molecules in *TRPML1*$^{-/-}$ cells. We propose that TRPML1 mediates a mechanism by which Fe^{2+} is released from late endosomes and lysosomes. Our results indicate that impaired iron transport may contribute to both haematological and degenerative symptoms of ML4 patients.

Release of Fe^{2+} from late endosomes and lysosomes into the cytosol is essential for cellular iron metabolism (see Supplementary Fig. 1)[6]. Because TRPML1 is ubiquitously expressed (in cells of every tissue)[5] and primarily localized in the late endosome and lysosome (LEL)[1,3,10], we propose that TRPML1 may act as an endolysosomal Fe^{2+} release channel. The intracellular localization of wild-type TRPML1 (refs 1, 3 and 10; see also Supplementary Fig. 2), however, makes it difficult to assay the function of this channel. To overcome this problem, we recently developed a functional assay for TRPML channels[11]. A spontaneous mouse mutation (A419P)[12] markedly increases TRPML3-mediated currents at the plasma membrane without altering its permeation properties[11]. Mice carrying this mutation (varitint-waddler, *Va*) are deaf and show skin pigmentation defects. When the *Va* mutation (proline substitution) was introduced into a homologous position in TRPML1 (V432P), the mutant channel $TRPML1^{Va}$ was (mis)localized in many cellular compartments including both LEL and the plasma membrane (Supplementary Fig. 2). Notably, $TRPML1^{Va}$, but not wild-type TRPML1, generated a cationic whole-cell current that can be measured by patch clamp[11].

We studied TRPML proteins by transiently expressing them in HEK293T cell lines. To monitor expression, TRPML channels were fused to enhanced green fluorescent protein (EGFP) at their amino termini. In response to a voltage step protocol, $TRPML1^{Va}$-expressing cells showed strong inwardly rectifying step currents (Fig. 1a and Supplementary Fig. 3) in a standard extracellular bath solution (a modified Tyrode's solution). $TRPML1^{Va}$-mediated current ($I_{TRPML1^{Va}}$) was 101 ± 8.6 pA pF^{-1} at -80 mV (mean $\pm$ s.e.m., $n = 48$). To mimic the acidic environments of lysosomes where the extracellular side (analogous to the intralysosomal luminal side) of the wild-type TRPML1 protein is exposed to, extracellular solutions were adjusted to pH 4.6. No sizable inward current was detected when *N*-methyl-D-glucamine ($NMDG^{+}$) was the only major cation in the bath (pH 4.6; Fig. 1b), indicating that $I_{TRPML1^{Va}}$ is a proton-impermeable cationic current[11]. Bath perfusion of 30 mM extracellular $[Fe^{2+}]$ ($[Fe^{2+}]_o$, pH 4.6) to $TRPML1^{Va}$-expressing cells induced large inwardly rectifying currents (Fig. 1c). The current density of $[Fe^{2+}]_o$-elicited $I_{TRPML1^{Va}}$ ($I_{Fe/TRPML1^{Va}}$) at -80 mV was 74 ± 8.6 pA pF^{-1} (mean $\pm$ s.e.m., $n = 24$). To keep Fe^{2+} in a reduced state, ascorbic acid was used as the counter anion in the bath solution. The only other main cation in the 30 mM Fe^{2+} solution was $NMDG^{+}$. Thus, $I_{Fe/TRPML1^{Va}}$ is carried solely by Fe^{2+}. Consistent with this, the amplitude of $I_{Fe/TRPML1^{Va}}$ was strongly dependent on $[Fe^{2+}]_o$ (Supplementary Fig. 4).

Both $I_{Fe/TRPML1^{Va}}$ (Fig. 1e) and $I_{TRPML1^{Va}}$[11] were enhanced at low pH. Compared to experiments conducted at physiological pH (7.4), $I_{Fe/TRPML1^{Va}}$ was enhanced ~2-fold at pH 4.6 (approximately the luminal pH of lysosomes). In cells transfected with $TRPML2^{Va}$ (a short splice variant; see Methods), a large current (23.0 ± 3.8 pA pF^{-1} at -80 mV, $n = 7$) was evoked by 30 mM Fe^{2+} (pH 4.6, Fig. 1e). Similar to $I_{TRPML1^{Va}}$, $I_{TRPML2^{Va}}$ was also strongly potentiated by low pH (Supplementary Fig. 5). The ratio of I_{Fe}/I_{Tyrode} for $TRPML2^{Va}$ was $139 \pm 22\%$ (at -80 mV, $n = 7$), which was even larger than the ratio for $TRPML1^{Va}$ ($72 \pm 7\%$, $n = 11$). In contrast with $TRPML1^{Va}$ and $TRPML2^{Va}$, the ratio of I_{Fe}/I_{Tyrode} for $TRPML3^{Va}$ was only $2.3 \pm 0.2\%$ ($n = 8$; Fig. 1f). Thus, Fe^{2+} permeability is specific for TRPML1 and TRPML2, but not for the closely related TRPML3. Cationic selectivity analysis suggested that $TRPML1^{Va}$ was also permeable to most other divalent trace metals including Mn^{2+} and Zn^{2+}, but not to Fe^{3+} (Supplementary Fig. 6). Detailed analyses of the permeation properties of $TRPML1^{Va}$ suggested that the conduction mechanism of Fe^{2+} resembled those of Na^{+} and Ca^{2+} (see Supplementary Figs 7–9).

[1]The Department of Molecular, Cellular, and Developmental Biology, The University of Michigan, 3089 Natural Science Building (Kraus), 830 North University, Ann Arbor, Michigan 48109, USA. [2]The Department of Cardiology, Children's Hospital Boston, Enders 1350, 320 Longwood Avenue, Boston, Massachusetts 02115, USA. [3]Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China. [4]Department of Pharmacology, Faculty of Health Science, University of Linköping, S-58185 Linköping, Sweden.

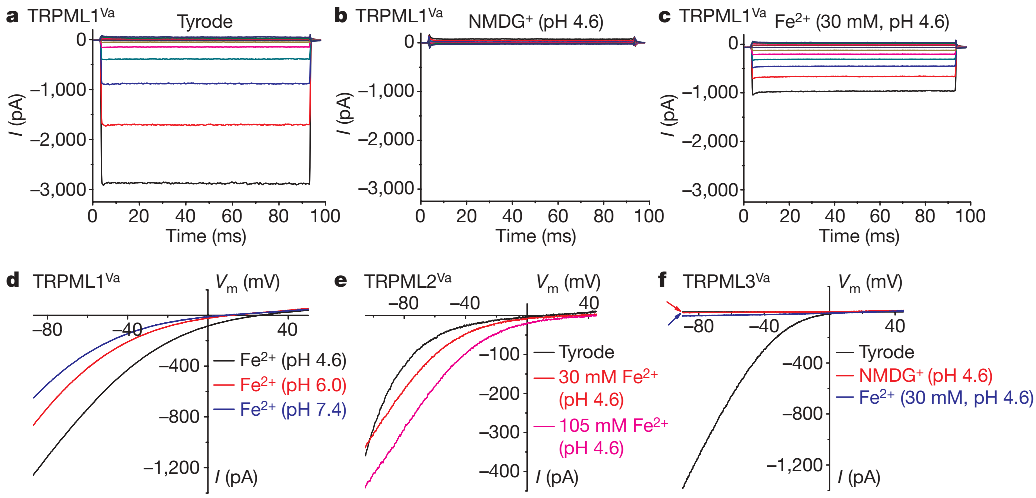


Figure 1 | TRPML1^{Va}-expressing cells have a constitutively active H$^+$-modulated Fe^{2+} current. a, A TRPML1^{Va}-expressing HEK293T cell showed a large inwardly rectifying whole-cell current elicited by voltage steps (from −140 mV to +80 mV in increments of 20 mV) in the standard extracellular (Tyrode's) bath solution. Step duration, 90 ms; holding potential, 0 mV; V_m, membrane potential. **b,** No significant inward current was detected in NMDG$^+$ (Na$^+$-free, Ca^{2+}-free, pH 4.6) solution. **c,** Inwardly rectifying step currents were evoked by 30 mM Fe^{2+} solution (pH 4.6) in the same cell as shown in **a** and **b**. **d,** pH-dependence of $I_{Fe/TRPML1^{Va}}$. Whole-cell currents were elicited by repeated voltage ramps (−100 to +100 mV; 400 ms) with a 4-s interval between ramps. Only a portion of the voltage protocol is shown. Holding potential, 0 mV. **e,** Large $I_{Fe/TRPML2^{Va}}$ was seen in the presence of 30 and 105 mM Fe^{2+} (pH 4.6). **f,** Little or no I_{Fe} was seen in a TRPML3^{Va}-expressing cell.

Next, we tested whether the Fe^{2+} permeability of TRPML1 is impaired by ML4 mutations. More than 15 ML4 mutations have been identified, most of which are located in the transmembrane regions of TRPML1 (Supplementary Fig. 10)[4,5,13,14]. We constructed six of these ML4 mutants into the TRPML1^{Va} background and investigated the Fe^{2+} permeability of these combined mutations. We found that all ML4 mutant TRPML1^{Va} channels had significantly smaller I_{Fe} than TRPML1^{Va}. In particular, T232P, D362Y and V446L mutations (combined with the *Va* mutation) completely eliminated I_{Fe}, as well as I_{Tyrode} (Fig. 2a and Supplementary Fig. 11), although the protein expression levels and subcellular localization pattern of these mutants were similar to those of the parental TRPML1^{Va} protein (Supplementary Figs 12 and 13). Patients carrying these mutations are reported to have severe phenotypes[5,15]. On the other hand, a large I_{Fe} was detected in F408Δ-TRPML1^{Va}-transfected cells. The average current density of this mutant, however, was still significantly smaller than $I_{Fe/TRPML1^{Va}}$. Patients carrying the F408Δ mutation have unusually mild phenotypes[13,15]. Small but evident I_{Fe} was detected with expression of two other mutants (R403C and F465L). A patient carrying the R403C mutation was reported to have a relatively mild phenotype[14]. Taken together, these results

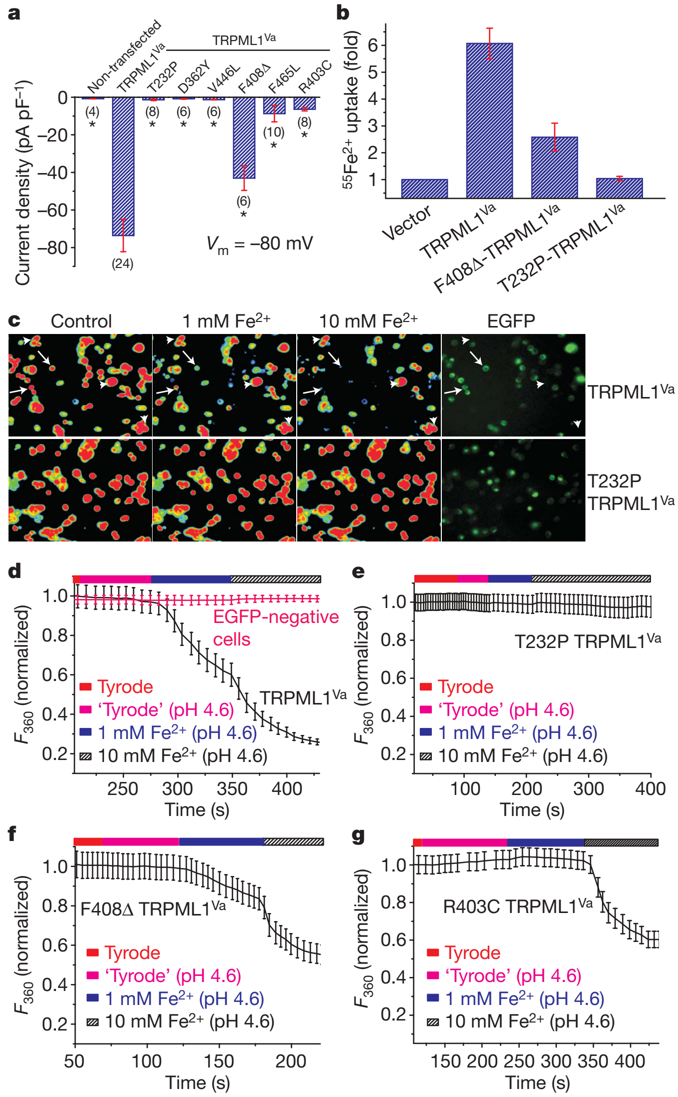


Figure 2 | Fe^{2+} permeability of the TRPML1 channel is impaired by ML4 mutations. a, Current densities (mean ± s.e.m., $n = 4$–10) of I_{Fe} (30 mM Fe^{2+}, pH 4.6) for TRPML1^{Va} and ML4 mutant TRPML1^{Va} channels. Asterisk indicates statistical difference ($P < 0.01$) compared to TRPML1^{Va}. **b,** ^{55}Fe^{2+} uptake (normalized) in HEK293T cells transfected with vector control, with TRPML1^{Va}, with F408Δ-TRPML1^{Va} and with T232P-TRPML1^{Va} constructs. Error bars indicate the standard deviation on the basis of two independent triplicate experiments. **c,** [Fe^{2+}]$_o$-dependent quenching of Fura-2 fluorescence in TRPML1^{Va}-transfected cells (arrows), but not in non-transfected control cells (arrowheads) or T232P-TRPML1^{Va}-transfected cells (bottom row). The fluorescence intensity was measured at an excitation wavelength of 360 nm (F_{360}). The original magnification used for all micrographs was ×200. **d,** Average normalized responses of EGFP-positive TRPML1^{Va}-transfected cells (typically $n = 20$–40 cells) to 1 or 10 mM Fe^{2+} (pH 4.6). **e,** No significant quenching was seen in T232P-TRPML1^{Va}-transfected cells. **f,** Slightly less quenching was observed for the F408Δ-TRPML1^{Va}-expressing cells. **g,** A small but significant quenching reaction was detected (with 10 mM Fe^{2+}) in R403C-TRPML1^{Va}-expressing cells.

indicate that ML4 mutations significantly impair the Fe^{2+} permeability of TRPML1. The degree of the impaired Fe^{2+} permeability seems to correlate well with disease severity observed in ML4 patients.

To investigate TRPML1^{Va}-mediated Fe^{2+} influx under less invasive conditions in intact cells, we measured $^{55}Fe^{2+}$ uptake at low pH (pH 5.8, 1 μM $^{55}Fe^{2+}$; Fig. 2b). Significant $^{55}Fe^{2+}$ uptake was seen in HEK293T cells transfected with TRPML1^{Va}, but not with T232P-TRPML1^{Va}. An intermediate $^{55}Fe^{2+}$ uptake was seen for F408Δ-TRPML1^{Va}. These results indicate that TRPML1^{Va} can mediate significant Fe^{2+} entry even at micromolar $[Fe^{2+}]_o$. We also adopted a fluorescence-based Fe^{2+} quenching assay to measure TRPML1^{Va}-mediated Fe^{2+} entry in HEK293T cells. Heavy metals such as Mn^{2+} and Fe^{2+} can bind Ca^{2+}-sensitive dyes such as Fura-2 with higher affinity, resulting in strong fluorescence quenching effects[16]. Substantial quenching of Fura-2 fluorescence was seen in TRPML1^{Va}-expressing cells after the addition of 1 mM Fe^{2+} (pH 4.6; Fig. 2d, f). Increased quenching was observed with the addition of higher concentrations of Fe^{2+} (10 mM, pH 4.6). In contrast, no significant quenching was detected in neighbouring non-transfected EGFP-negative cells or TRPML3^{Va}-transfected cells (1 mM Fe^{2+}; Supplementary Fig. 14). ML4 mutant TRPML1^{Va}-expressing HEK293T cells showed an impairment of Fe^{2+} quenching. No significant Fe^{2+} quenching was seen in T232P- (Fig. 2e), D362Y-, or V446L-TRPML1^{Va}-expressing cells (data not shown). Slightly less quenching was observed for F408Δ-TRPML1^{Va} (Fig. 2f). In contrast, R403C-TRPML1^{Va} (Fig. 2g) and F465L-TRPML1^{Va} showed no response to the addition of 1 mM Fe^{2+}. However, a higher concentration of Fe^{2+} (10 mM) induced a slow but significant quenching. These results agree with the electrophysiological measurements of ML4 mutants, as well as the disease severity of ML4 patients carrying

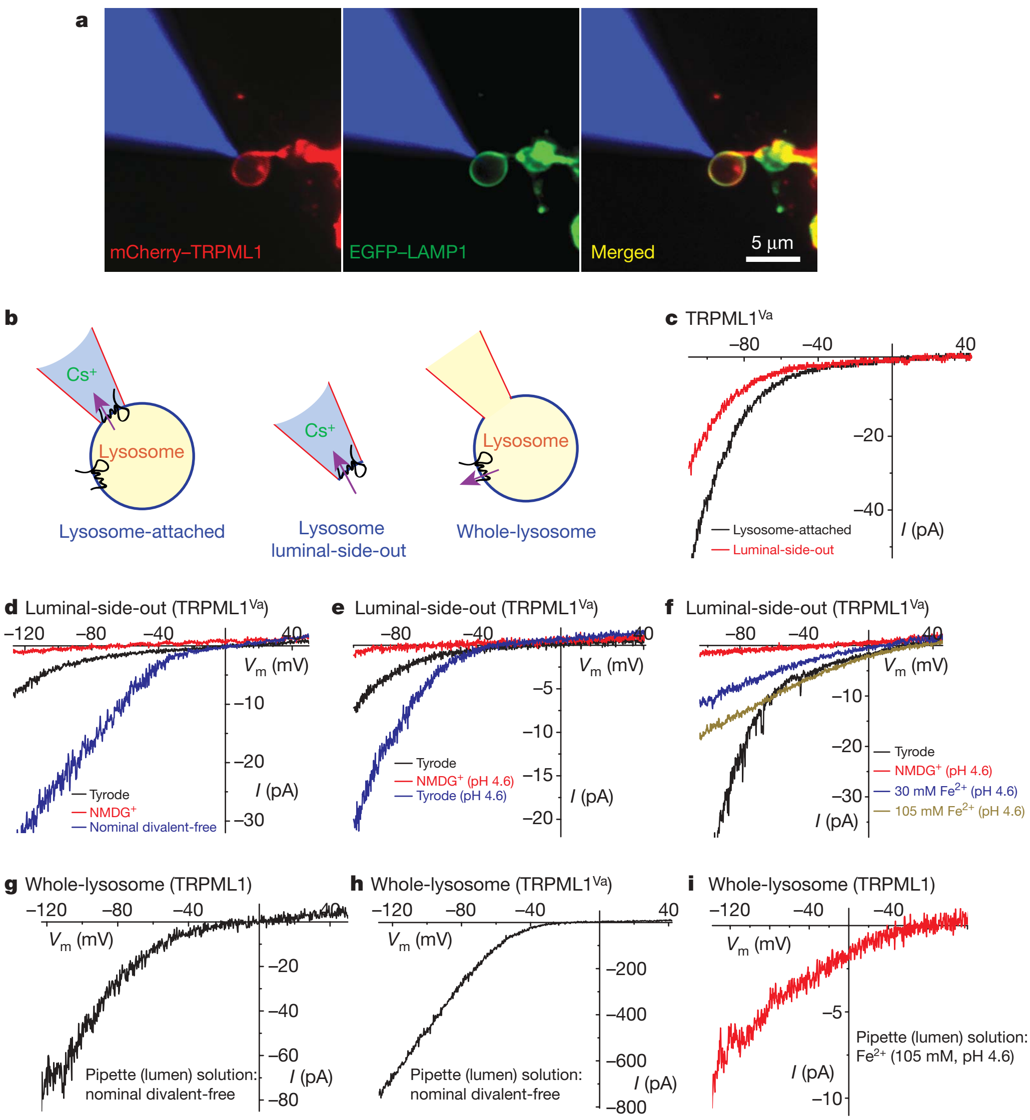


Figure 3 | TRPML1 conducts Fe^{2+} in late endosomes and lysosomes. **a**, Co-localization of mCherry–TRPML1 and EGFP–LAMP1 at the membrane of an isolated enlarged LEL (see Methods). The patch pipette was filled with red rhodamine B dye (shown in blue for the purpose of illustration). **b**, Cartoons of three distinct patch-clamp configurations of lysosomal recordings: lysosome-attached, lysosome luminal-side-out and whole-lysosome. In each configuration, the pink arrow indicates the direction of the inward (at negative potentials; flow out of the lysosomes) current mediated by TRPML1 (as shown in **c**–**i**). **c**, Lysosomal $I_{TRPML1^{Va}}$. Switching from lysosome-attached to (lysosome) luminal-side-out configuration significantly reduced the amplitude of the current. The luminal-side-out patch was exposed to the Tyrode's solution. A Cs^+-based solution (147 mM Cs-methanesulphonate (Cs-MSA)) was used as a pipette solution for both configurations. **d**, $NMDG^+$-impermeable lysosomal $I_{TRPML1^{Va}}$ was much larger in the absence of divalent cations (nominal divalent-free). **e**, Lowering pH potentiated lysosomal $I_{TRPML1^{Va}}$. **f**, $I_{Fe/TRPML1^{Va}}$ induced by 30 mM and 105 mM Fe^{2+}. **g**, Whole-lysosome current in an enlarged lysosome expressing wild-type TRPML1. The pipette (lumen) solution contained nominal divalent-free Tyrode solution. A Cs^+-based bath solution (147 mM Cs-MSA) was used. **h**, Whole-lysosome $I_{TRPML1^{Va}}$. **i**, Whole-lysosome $I_{Fe/TRPML1}$. The pipette (lumen) solution contained 105 mM Fe^{2+} (pH 4.6).

these mutations[5,13–15]. Thus, ML4 phenotypes result from a loss-of-function of TRPML1 with respect to Fe^{2+} and/or Ca^{2+} permeability.

To validate our use of TRPML1^{Va} as a model to extrapolate potential intracellular functions of wild-type TRPML1, and more importantly, to confirm whether wild-type TRPML1 conducts Fe^{2+} from the lumen of LEL into the cytosol, we performed patch-clamp recordings directly on native LEL membranes (Fig. 3a and Supplementary Fig. 15; see Methods). In TRPML1^{Va}-positive enlarged LEL, large inwardly rectifying currents were seen under the lysosome-attached configuration (Fig. 3b, c). The currents became smaller when the patch was excised (lysosome luminal-side-out) and exposed to Tyrode's solution (Fig. 3c). These large inwardly rectifying currents were not seen in TRPML1^{Va}-negative vacuoles, suggesting that these lysosomal currents are mediated by TRPML1^{Va} (hereafter referred to as lysosomal $I_{TRPML1^{Va}}$). Note that inward lysosomal currents are actually currents that flow out of the lysosomes. Similar to $I_{TRPML1^{Va}}$ recorded at the plasma membrane, lysosomal $I_{TRPML1^{Va}}$ was impermeable to $NMDG^+$ and H^+, but was potentiated by low pH or by removal of divalent cations in the bath solution (nominal divalent-free; Fig. 3d, e). When the luminal-side-out patches were exposed to 30 mM or 105 mM Fe^{2+} solutions (pH 4.6), smaller and less rectifying currents were seen, resembling $I_{Fe/TRPML1^{Va}}$ at the plasma membrane. These results indicate that TRPML1^{Va} behaves similarly at both plasma and lysosomal membranes. Likewise, ML4 mutant TRPML1^{Va} channels also behaved similarly as their plasma membrane counterparts (Supplementary Fig. 16), consistent with our conclusion that ML4 mutations primarily affect the channel function of TRPML1.

Next we performed whole-lysosome recordings on enlarged LEL expressing wild-type TRPML1 and TRPML1^{Va}. Significant inwardly rectifying currents were seen in wild-type TRPML1-positive LEL (81.0 ± 9.0 pA pF^{-1} at -120 mV, $n = 6$; Fig. 3g) with nominal divalent-free solution in the luminal side (that is, pipette solution), although the current amplitude is still one order of magnitude smaller than the whole-lysosome current recorded from TRPML1^{Va}-positive LEL ($1{,}713 \pm 404$ pA pF^{-1} at -120 mV, $n = 4$; Fig. 3h). Because we were not able to record any significant whole-cell current from wild-type TRPML1-expressing HEK293T cells[11], our results indicate that the *Va* mutation affects both channel gating and trafficking (between LEL and the plasma membrane) of TRPML1. When isotonic (105 mM) Fe^{2+} solution was included in the luminal side of wild-type TRPML1-positive enlarged LEL, inwardly rectifying currents with very positive reversal potential were observed (18 ± 4 pA pF^{-1} at -120 mV, $n = 4$; Fig. 3i). Collectively, these results indicate that wild-type TRPML1 is a lysosomal Fe^{2+}-permeable channel, and that the *Va* mutation does not alter the permeation properties of the TRPML1 channel.

Retention of Fe^{2+} in LEL due to loss or inactivation of the Fe^{2+} release mechanism may result in an insufficient supply of cytosolic Fe^{2+}. We therefore measured the levels of free (chelatable) intracellular Fe^{2+} in skin fibroblasts from a ML4 patient (*TRPML1*$^{-/-}$) and the non-diseased parent (*TRPML1*$^{+/-}$) using a fluorescence-based iron de-quenching imaging method (see Methods)[17]. Chelatable Fe^{2+} levels were significantly lower in *TRPML1*$^{-/-}$ cells compared to the control *TRPML1*$^{+/-}$ cells (Fig. 4a, b). TRPML1-deficient skin fibroblast cells show autofluorescence (Fig. 4c) at excitation wavelengths between 440 nm and 500 nm, reminiscent of accumulated lipofuscin-like substances previously reported[18]. The autofluorescence observed was primarily localized in LAMP1 (a marker for LEL)-positive compartments (Fig. 4c), indicative of a lysosomal dysfunction in *TRPML1*$^{-/-}$ cells. Lipofuscin, also referred to as an 'ageing pigment', is a non-degradable oxidized polymeric substance containing proteins, lipids, carbohydrates and iron[19]. The production of lipofuscin is facilitated by increased intralysosomal Fe^{2+} levels[19]. Our results indicate that an accumulation of autofluorescent lipofuscin-like molecules in *TRPML1*$^{-/-}$ cells might result from impaired intralysosomal iron metabolism. Consistent with this, lysosomal iron staining experiments showed that *TRPML1*$^{-/-}$ cells had higher lysosomal iron content than control cells (Supplementary Fig. 17).

We have demonstrated that TRPML1 has a critical role in cellular iron homeostasis by showing the cytosolic Fe^{2+} deficiency with concurrent intralysosomal iron overload in ML4 cells. ML4 cells show a defect in the late endocytic pathway[10,20,21], suggesting that TRPML1 may be required for Ca^{2+}-dependent fusion or fission of LEL. However, this defect in vesicular transport cannot explain the cytosolic Fe^{2+} deficiency that we observed because the internalization and recycling of the transferrin receptor is normal in ML4 cells[20], and lysosomal Fe^{2+} release (into the cytosol) is probably mediated directly by an iron-conducting channel/transporter, as is the case for DMT1 (ref. 22). Therefore, by far the simplest explanation of our results is that the Fe^{2+} release pathway between the cytosol and the lysosome lumen is blocked or inactivated in cells with ML4 mutations. Although a ML4-induced loss of Ca^{2+} permeability may result in a defect in lysosomal trafficking and subsequent accumulation of various lipids in LEL, the degradation of these materials is normal[20,21]. Therefore, these accumulated substances would become most harmful only if they are oxidized to generate lipofuscin-like non-degradable materials in the presence of high intralysosomal Fe^{2+} (ref. 19). Lipofuscin-like molecules preferentially accumulate in post-mitotic cells such as muscle cells and neurons, which are primarily affected in ML4 patients[21]. Our work suggests that lysosome-targeting iron chelators might alleviate the degenerative symptoms of ML4 patients.

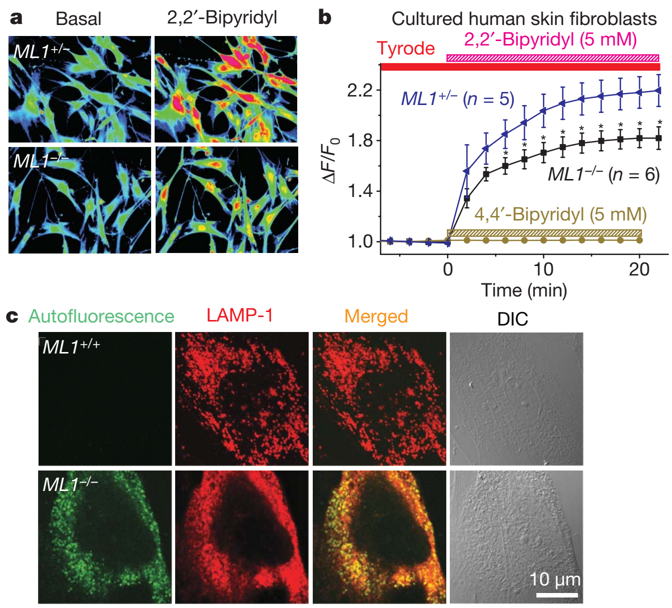


Figure 4 | TRPML1-deficient cells have reduced free (chelatable) iron levels and lysosomal autofluorescence. a, Cultured *TRPML1*$^{-/-}$ (*ML1*$^{-/-}$) skin fibroblasts showed less de-quenching of the iron-sensitive fluorescence than *ML1*$^{+/-}$ cells did. De-quenching was achieved by preloading the fibroblasts with an iron-sensitive dye, Phen Green SK (PG SK), and then adding the membrane-permeable transition metal chelator, 2,2′-bipyridyl (BPD). 2,2′-BPD is predicted to chelate free cellular iron (also referred to as chelatable or labile iron), which subsequently increases PG SK fluorescence. The original magnification used was ×200. **b,** The average 2,2′-BPD induced normalized change of fluorescence ($\Delta F/F_0$) for ML4 cells (*ML1*$^{-/-}$; $n = 6$ experiments) is significantly (asterisk, $P < 0.01$) lower than for the parental *ML1*$^{+/-}$ cells. Fibroblast cells ($n = 10$–20) were analysed for each individual experiment. 4,4′-BPD, a 2,2′-BPD analogue that cannot bind Fe^{2+}, did not induce any significant restoration of PG SK fluorescence. Error bars, s.e.m. **c,** An *ML1*$^{-/-}$ skin fibroblast cell showed autofluorescence in LAMP1-positive compartments. Autofluorescence (green) was detected within a range of excitation wavelengths (shown with excitation at 480 nm). No significant autofluorescence was observed for a *ML1*$^{+/+}$ cell. Lysosomes were stained with a LAMP1 antibody (red). Differential interference contrast (DIC) images are shown for comparison.

From the cell biology perspective, an important question to address is how TRPML channels are regulated by various cytosolic and luminal factors, and/or by proteins and lipids in the LEL membrane, especially those of which are known to be involved in endolysosomal trafficking.

METHODS SUMMARY

Endolysosomal electrophysiology. HEK293T cells were either transfected with EGFP–TRPML1 alone or co-transfected with mCherry–TRPML1 and EGFP–LAMP1 (a marker for LEL). The size of the LEL is usually $<0.5\ \mu m$, which is suboptimal for patch clamping. We therefore treated cells with $1\ \mu M$ vacuolin-1 (for 1–2 h), a small chemical known to increase the size of endosomes and lysosomes selectively[23]. Large vacuoles (up to $3\ \mu m$; capacitance $= 321 \pm 101$ fF, $n = 11$) were observed in most vacuolin-treated cells (Supplementary Fig. 15). Occasionally, enlarged LEL could also been obtained from TRPML1-transfected cells without vacuolin-1 treatment. No significant difference in TRPML channel properties was seen for enlarged LEL with or without vacuolin-1 treatment. The vacuoles that were positive for both mCherry–TRPML1 and EGFP–LAMP1 were considered as enlarged LEL. Whole-lysosome, lysosome-attached and lysosome luminal-side-out recordings were performed on isolated enlarged LEL, similar to what was performed in enlarged endosomes[24]. In brief, a patch pipette (electrode) was pressed against a cell and then quickly pulled away from the cell to slice the cell membrane. Enlarged LEL were released into the dish and identified by monitoring the TRPML1–EGFP, the TRPML1–mCherry or the LAMP1–EGFP fluorescence.

Iron quenching imaging. HEK293T cells were loaded with $5\ \mu M$ Fura-2 AM (Molecular Probes) in culture medium at 37 °C for 60 min. Cells were washed in Tyrode's solution for 10–30 min and fluorescence intensity at wavelength 360 nm (F_{360}) was recorded on an EasyRatioPro system (Photon Technology International). EGFP-positive cells were identified by monitoring fluorescence intensity at wavelength 470 nm (F_{470}). Images were analysed using ERP software.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 March; accepted 1 August 2008.
Published online 14 September 2008.

1. Venkatachalam, K., Hofmann, T. & Montell, C. Lysosomal localization of TRPML3 depends on TRPML2 and the mucolipidosis-associated protein TRPML1. *J. Biol. Chem.* **281,** 17517–17527 (2006).
2. Clapham, D. E. TRP channels as cellular sensors. *Nature* **426,** 517–524 (2003).
3. Nilius, B., Owsianik, G., Voets, T. & Peters, J. A. Transient receptor potential cation channels in disease. *Physiol. Rev.* **87,** 165–217 (2007).
4. Bassi, M. T. *et al.* Cloning of the gene encoding a novel integral membrane protein, mucolipidin-and identification of the two major founder mutations causing mucolipidosis type IV. *Am. J. Hum. Genet.* **67,** 1110–1120 (2000).
5. Sun, M. *et al.* Mucolipidosis type IV is caused by mutations in a gene encoding a novel transient receptor potential channel. *Hum. Mol. Genet.* **9,** 2471–2478 (2000).
6. Hentze, M. W., Muckenthaler, M. U. & Andrews, N. C. Balancing acts: molecular control of mammalian iron metabolism. *Cell* **117,** 285–297 (2004).
7. Lee, D. W., Andersen, J. K. & Kaur, D. Iron dysregulation and neurodegeneration: the molecular connection. *Mol. Interv.* **6,** 89–97 (2006).
8. Kidane, T. Z., Sauble, E. & Linder, M. C. Release of iron from ferritin requires lysosomal activity. *Am. J. Physiol. Cell Physiol.* **291,** C445–C455 (2006).
9. Gunshin, H. *et al.* Slc11a2 is required for intestinal iron absorption and erythropoiesis but dispensable in placenta and liver. *J. Clin. Invest.* **115,** 1258–1266 (2005).
10. Pryor, P. R., Reimann, F., Gribble, F. M. & Luzio, J. P. Mucolipin-1 is a lysosomal membrane protein required for intracellular lactosylceramide traffic. *Traffic* **7,** 1388–1398 (2006).
11. Xu, H., Delling, M., Li, L., Dong, X. & Clapham, D. E. Activating mutation in a mucolipin transient receptor potential channel leads to melanocyte loss in varitint-waddler mice. *Proc. Natl Acad. Sci. USA* **104,** 18321–18326 (2007).
12. Di Palma, F. *et al.* Mutations in *Mcoln3* associated with deafness and pigmentation defects in varitint-waddler (*Va*) mice. *Proc. Natl Acad. Sci. USA* **99,** 14994–14999 (2002).
13. Altarescu, G. *et al.* The neurogenetics of mucolipidosis type IV. *Neurology* **59,** 306–313 (2002).
14. Goldin, E. *et al.* Transfer of a mitochondrial DNA fragment to *MCOLN1* causes an inherited case of mucolipidosis IV. *Hum. Mutat.* **24,** 460–465 (2004).
15. Bargal, R., Goebel, H. H., Latta, E. & Bach, G. Mucolipidosis IV: novel mutation and diverse ultrastructural spectrum in the skin. *Neuropediatrics* **33,** 199–202 (2002).
16. Kress, G. J., Dineley, K. E. & Reynolds, I. J. The relationship between intracellular free iron and cell injury in cultured neurons, astrocytes, and oligodendrocytes. *J. Neurosci.* **22,** 5848–5855 (2002).
17. Petrat, F., de Groot, H. & Rauen, U. Determination of the chelatable iron pool of single intact cells by laser scanning microscopy. *Arch. Biochem. Biophys.* **376,** 74–81 (2000).
18. Goldin, E., Blanchette-Mackie, E. J., Dwyer, N. K., Pentchev, P. G. & Brady, R. O. Cultured skin fibroblasts derived from patients with mucolipidosis 4 are autofluorescent. *Pediatr. Res.* **37,** 687–692 (1995).
19. Kurz, T., Terman, A., Gustafsson, B. & Brunk, U. T. Lysosomes in iron metabolism, ageing and apoptosis. *Histochem. Cell Biol.* **129,** 389–406 (2008).
20. Chen, C. S., Bach, G. & Pagano, R. E. Abnormal transport along the lysosomal pathway in mucolipidosis, type IV disease. *Proc. Natl Acad. Sci. USA* **95,** 6373–6378 (1998).
21. Zeevi, D. A., Frumkin, A. & Bach, G. TRPML and lysosomal function. *Biochim. Biophys. Acta.* **1772,** 851–858 (2007).
22. Andrews, N. C. & Schmidt, P. J. Iron homeostasis. *Annu. Rev. Physiol.* **69,** 69–85 (2007).
23. Cerny, J. *et al.* The small chemical vacuolin-1 inhibits Ca^{2+}-dependent lysosomal exocytosis but not cell resealing. *EMBO Rep.* **5,** 883–888 (2004).
24. Saito, M., Hanson, P. I. & Schlesinger, P. Luminal chloride-dependent activation of endosome calcium channels: patch clamp study of enlarged endosomes. *J. Biol. Chem.* **282,** 27327–27333 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work is supported by start-up funds to H.X. from the Department of Molecular, Cellular, and Developmental Biology and Biological Science Scholar Program, University of Michigan. We thank U. Brunk, M. Saito, R. Hume, C. Duan, M. Akaaboune, J. Kuwada, S. Low, S. Punthambaker and S. Dellal for assistance, and D. Clapham, N. Andrews, L. DeFelice, L. Yue, D. Ren, C. Jiang and S. Xu for comments on an earlier version of the manuscript. We also thank K. Kiselyov for sharing his unpublished results on lysosomal iron staining of ML4 cells. We appreciate the encouragement and helpful comments from other members of the Xu laboratory.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to H.X. (haoxingx@umich.edu).

METHODS

Molecular biology and biochemistry. Full-length mouse TRPML1 and TRPML3, and a short splicing variant of mouse TRPML2 (GeneBank accession NP_001005846), were cloned into the EGFP-C2 vector (Clontech), the mCherry vector, or haemagglutinin-tag containing pcDNA3 as described previously[11]. No significant functional difference was observed among EGFP-tagged, mCherry-tagged and haemagglutinin-tagged $TRPML^{Va}$ constructs. ML4 mutants were constructed on a $TRPML1^{Va}$ background using a site-directed mutagenesis kit (Qiagen). All constructs were confirmed by sequencing analysis and protein expression was verified by western blotting. HEK293T cells were transiently transfected with wild-type TRPML1, $TRPML1^{Va}$ and ML4 mutants for electrophysiology, confocal imaging, $^{55}Fe^{2+}$ uptake and Fe^{2+} quenching assays. TRPML1 western blot analyses were performed with an anti-EGFP monoclonal antibody (Covance).

Confocal imaging. All images were taken using a Leica (TCS SP5) confocal microscope. LAMP1 antibody was from the Iowa Hybridoma Bank and LysoTracker was from Molecular Probes (Invitrogen).

Mammalian cell electrophysiology. Recordings were performed in transiently transfected HEK293T cells. Cells were cultured at 37 °C, transfected using Lipofectamine 2000 (Invitrogen) and plated onto glass coverslips. No significant difference was observed for EGFP-tagged versus untagged $TRPML1^{Va}$-transfected cells. Unless otherwise stated, pipette solution contained 147 mM Cs, 120 mM methanesulphonate (MSA), 4 mM NaCl, 10 mM EGTA, 2 mM Na_2-ATP, 2 mM $MgCl_2$, 20 mM HEPES (pH 7.2; free $[Ca^{2+}]_i$ <10 nM). Standard extracellular bath solution (modified Tyrode's solution) contained 153 mM NaCl, 5 mM KCl, 2 mM $CaCl_2$, 1 mM $MgCl_2$, 20 mM HEPES and 10 mM glucose (pH 7.4). To reduce the background from an endogenous Cl^- conductance activated by protons that was strongly outwardly rectifying[11,25], gluconate$^-$ or MSA^- was used to replace most of the Cl^- (remaining $[Cl^-]_o$ = 5–10 mM) for all low-pH bath solutions. Low-pH 'Tyrode's' solution contained 150 mM Na-gluconate, 5 mM KCl, 2 mM $CaCl_2$, 1 mM $MgCl_2$, 10 mM glucose, 10 mM HEPES and 10 mM MES (pH 4.6). $NMDG^+$ solution contained 160 mM NMDG, 20 mM HEPES, 10 mM glucose (pH 7.4). Low-pH $NMDG^+$ solution contained 150 mM NMDG, 10 mM glucose, 10 mM MES, 10 HEPES (pH adjusted to 4.6 using MSA). 'Isotonic' metal solutions contained 105 mM metal ions, 30 mM glucose, 10 mM HEPES, 10 mM MSA and 0–30 mM $NMDG^+$ (pH 4.6). Further metal solutions (1, 3, 10 and 30 mM) were derived from mixing isotonic solutions with low-pH $NMDG^+$ solutions at various ratios. Monovalent (nominal divalent-free) solutions contained 10 mM glucose, 20 mM HEPES, 160 mM NaCl and 5 mM KCl (pH 7.4; free Ca^{2+} <1–10 μM). All solutions were applied by a fast perfusion system to achieve a complete solution exchange within a few seconds. Data were collected using an Axopatch 2A patch clamp amplifier, Digidata 1440 and pClamp 10.0 software (Axon Instruments). Whole-cell currents and single channel recordings were digitized at 10 kHz and filtered at 2 kHz. Capacity current was reduced as much as possible using the amplifier circuitry. Series resistance compensation was 60%–85%. All experiments were conducted at room temperature (~21–23 °C) and all recordings were analysed with pCLAMP10 (Axon Instruments) and Origin 7.5 (Origin Lab).

Iron de-quenching imaging. Cultured skin fibroblast cells from a ML4 patient (*TRPML1*$^{-/-}$, clone WG0909) and corresponding heterozygous parent (*TRPML1*$^{+/-}$, clone WG0987) were obtained from the Repository for Mutant Human Cell Strains of Montreal Children's Hospital (Montreal, Canada). Fibroblast cells were loaded with 20 μM Phen Green SK (Molecular Probes) in culture medium at 37 °C for 20 min followed by a 15 min wash. Cellular fluorescence was excited at 488 nm on the PTI ERP system. Quenching of fluorescence was induced by the addition of the membrane-permeable transition metal chelator 2,2′-bipyridyl (5 mM)[17]. The 2,2′-bipyridyl analogue 4,4′-bipyridyl, which cannot bind or chelate Fe^{2+}, was used as a control. Because many variables such as dye loading may contribute to the variation of basal fluorescence (F) of PG SK, the normalized change of fluorescence ($\Delta F/F$) was used as readout to estimate the change in cytosolic Fe^{2+} levels.

Iron uptake assay. Iron uptake was measured as described previously[25]. In brief, 30 h after transfection, HEK293T cells were washed twice with prewarmed PBS and then incubated with 1 ml of prewarmed uptake assay buffer (25 mM Tris, 25 mM MES, 140 mM NaCl, 5.4 mM KCl, 5 mM glucose, 1.8 mM $CaCl_2$, 800 μM MgSO4, 50 μM ascorbic acid, pH 5.8) containing ^{55}Fe-nitrilotriacetic acid (NTA, 1 μM ^{55}Fe) at 37 °C for 5 min. ^{55}Fe-NTA was made by mixing $^{55}FeCl_3$ (PerkinElmer) with 100 mM NTA in a 1:50 molar ratio. Ascorbic acid (100 mM) was used to promote the formation and maintenance of ferrous (Fe^{2+}) iron (adjusted to pH 5.8). All uptake assays were performed with 20 μM Fe^{2+} (with a 1:20 molar ratio for $^{55}FeCl_3$ and $FeSO_4$) at pH 5.8. Assays were stopped by the addition of 2 ml ice-cold PBS. After washing three times with ice-cold PBS, the cells were detached and collected by adding 1 ml 0.25% trypsin-EDTA. Parallel experiments were conducted at 0 °C to measure the cell surface ^{55}Fe binding, which was subtracted from the values at 37 °C to obtain the net iron uptake. Cell-associated radioactivity was measured with liquid scintillation spectrometry.

Lysosomal iron staining. Lysosome iron was stained using a modified sulphide-silver method[19]. In brief, fibroblast cells were grown on coverslips and exposed to 50 μM Fe^{3+}-dextran overnight. Cells were then washed with PBS and fixed with 2% glutaraldehyde in 0.1 M Na-cacodylate buffer with 0.1 M sucrose (pH 7.2) for 1.5 h at room temperature (22 °C). Next cells were sulphidated at ~pH 9 with 1% (w/v) ammonium sulphide in 70% (v/v) ethanol for 15 min. The development was then performed using a physical, colloid-protected developer containing silver-lactate and hydroquinone for 1–2 h.

Reagents. 2,2'-BPD, 4,4'-BPD, iron ascorbate, ferric chloride, iron dextran, and all other metal salts were from Sigma Chemicals.

Data analysis. Most data are presented as the mean ± s.e.m. Statistical comparisons were made using analysis of variance (ANOVA). A P value < 0.05 was considered statistically significant.

25. Xu, H., Jin, J., DeFelice, L. J., Andrews, N. C. & Clapham, D. E. A spontaneous, recurrent mutation in divalent metal transporter-1 exposes a calcium entry pathway. *PLoS Biol.* **2**, E50 (2004).

Vol 455 | 16 October 2008 | doi:10.1038/nature07325

LETTERS

Promoter-driven splicing regulation in fission yeast

Alberto Moldón[1], Jordi Malapeira[1], Natalia Gabrielli[1], Madelaine Gogol[2], Blanca Gómez-Escoda[1], Tsvetomira Ivanova[1], Chris Seidel[2] & José Ayté[1]

The meiotic cell cycle is modified from the mitotic cell cycle by having a pre-meiotic S phase that leads to high levels of recombination, two rounds of nuclear division with no intervening DNA synthesis and a reductional pattern of chromosome segregation. Rem1 is a cyclin that is only expressed during meiosis in the fission yeast *Schizosaccharomyces pombe*. Cells in which *rem1* has been deleted show decreased intragenic meiotic recombination and a delay at the onset of meiosis I (ref. 1). When ectopically expressed in mitotically growing cells, Rem1 induces a G1 arrest followed by severe mitotic catastrophes. Here we show that *rem1* expression is regulated at the level of both transcription and splicing, encoding two proteins with different functions depending on the intron retention. We have determined that the regulation of *rem1* splicing is not dependent on any transcribed region of the gene. Furthermore, when the *rem1* promoter is fused to other intron-containing genes, the chimaeras show a meiotic-specific regulation of splicing, exactly the same as endogenous *rem1*. This regulation is dependent on two transcription factors of the forkhead family, Mei4 (ref. 2) and Fkh2 (ref. 3). Whereas Mei4 induces both transcription and splicing of *rem1*, Fkh2 is responsible for the intron retention of the transcript during vegetative growth and the pre-meiotic S phase.

The switch from mitosis to meiosis is accompanied by a deep change in gene expression profiles[4], as shown by genome-wide analyses in both *S. pombe*[5] and *Saccharomyces cerevisiae*[6,7]. One of the genes that is expressed and differentially spliced during meiosis is *rem1* (refs 1 and 8). Δ*rem1* cells complete meiosis and produce four-spore asci with a viability similar to wild-type cells (>90%). However, meiotic intragenic recombination is reduced twofold to threefold[1]. Notably, *rem1* splicing takes place just after the recombination events, pointing to the possibility that full-length Rem1 might not be involved in the regulation of recombination. To determine whether this is the case, we measured recombination of a Δ*rem1* strain transformed with different plasmids and compared it with a wild-type strain. As shown in Fig. 1a, in the absence of *rem1* recombination efficiency was decreased twofold; it was restored with a plasmid containing the genomic region of *rem1* but not with an empty vector. Normal levels of recombination were not restored when the plasmid contained a mutation in the first codon of *rem1* (ATG-less), indicating that *rem1* messenger RNA is not directly involved in the regulation of recombination. However, a plasmid containing a mutation in the splicing donor site that produces an mRNA that is never spliced (NoS; Fig. 1b and Supplementary Fig. 1) and, as a consequence of the presence of a stop codon in the intron, renders a peptide encoded exclusively by the first exon, can restore wild-type levels of recombination. Thus, the 17 kDa peptide encoded by the first exon is sufficient to bring about normal levels of recombination, whereas the second exon (which contains the cyclin box) is dispensable for the recombination function of Rem1 (Fig. 1c). Thus, the timely regulation of splicing indicates that there are different functions of *rem1*.

To study this regulation, we analysed splicing of different chimaeras of *rem1* during meiosis (Fig. 2). Figure 2a shows the splicing profile of the wild-type *rem1*, compared to a mutant in which the 3′

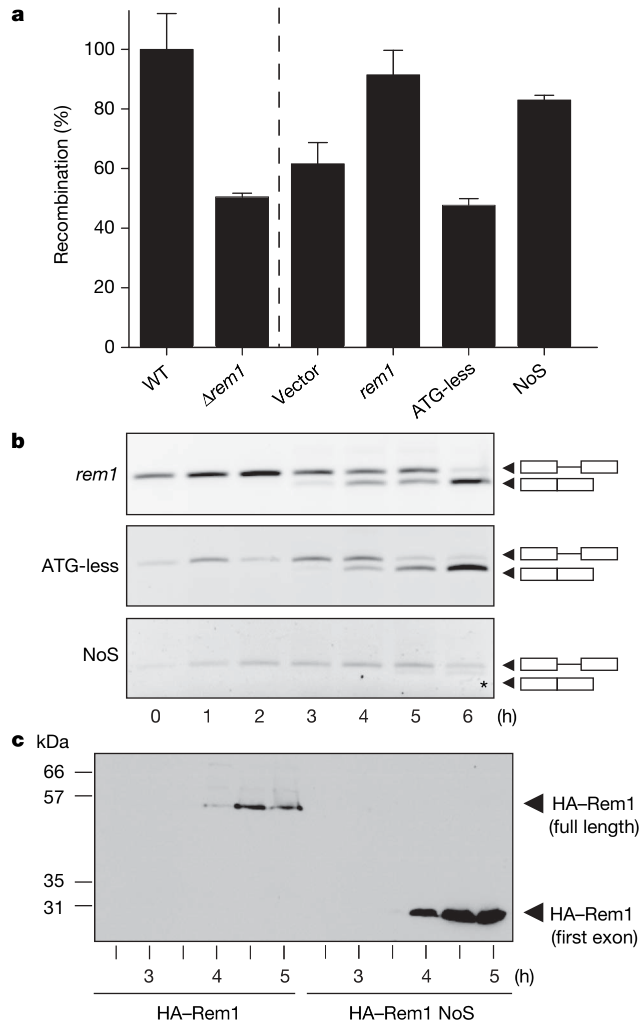


Figure 1 | Alternatively spliced *rem1* restores normal recombination levels. **a**, Recombination was measured in wild-type (WT) and Δ*rem1* strains, and in Δ*rem1* strains transformed with different constructs (right of dashed line). Mean and s.d. are shown; $n = 5$. **b**, *rem1* splicing was monitored by RT–PCR in cells expressing wild-type *rem1* and the mutants ATG-less and NoS. The asterisk marks the splicing product from a cryptic 5′ splicing. **c**, Protein extracts were prepared from meiotic cells expressing wild-type Rem1 (haemagglutinin (HA)–Rem1) or the NoS mutant (HA–Rem1 NoS) and proteins were detected using anti-HA antibodies. Molecular mass markers are indicated on the left.

[1]Oxidative Stress and Cell Cycle Group, Universitat Pompeu Fabra, C/Doctor Aiguader 88, Barcelona 08003, Spain. [2]Stowers Institute, Kansas City, Missouri 64110, USA.

untranslated region (UTR) and most of the second exon had been deleted. During synchronous meiosis (Supplementary Fig. 2), both constructs showed a wild-type profile of splicing with minor differences at the end of meiosis, probably due to the lack of a polyadenylation site in the mutant affecting its stability. This contrasted with the profile detected when the promoter was changed. As shown in Fig. 2b, when *rem1* expression was under the control of the inducible *nmt41* promoter (at the *rem1* locus), splicing was constitutive, either with or without the 5′ UTR from *rem1*. Furthermore, when *rem1* was under the control of the *res2* promoter (another intron-containing gene), splicing was also constitutive (Fig. 2c). To ascertain fully whether the promoter was responsible for the regulation of the splicing, we proceeded with a reverse strategy. We fused the promoter of *rem1*, either with (data not shown) or without the *rem1* 5′ UTR region, to different reporter genes such as the meiosis-specific *rec6* (data not shown) and the ubiquitous *cdc2*. Splicing of the different chimaeras as well as the endogenous genes was followed by polymerase chain reaction with reverse transcription (RT–PCR), as cells progressed into synchronous meiosis (Supplementary Fig. 3). As shown in Fig. 2d, the *rem1* promoter is sufficient to regulate intron retention during vegetative growth and the early stages of meiosis (top right panel). Furthermore, this splicing regulation is dependent on the presence of Mei4 (see later), because intron retention of the chimaeras can be detected in Δ*mei4* cells during the entire time course, as occurs with endogenous *rem1* (Fig. 2d; compare Δ*mei4* in top left and right panels). Conversely, endogenous *cdc2* (bottom panel) shows Mei4-independent constitutive splicing. Thus, the promoter of *rem1* is sufficient to direct intron retention in genes with up to four introns, such as *cdc2*.

The forkhead family is highly conserved in eukaryotes with roles in cell cycle control, cell death and differentiation[9,10]. FLEX boxes are the *cis*-elements recognized by the forkhead transcription factors[2]. There are two putative FLEX boxes in the *rem1* promoter, as indicated in Supplementary Fig. 4a. In fission yeast, four genes code for forkhead proteins: *sep1* (ref. 11), *fhl1* (ref. 12), *mei4* (ref. 2) and *fkh2* (refs 3 and 13). We have previously shown that Mei4 is required for proper regulation of *rem1* transcription[1]. Mei4 overexpression in mitotically growing cells can induce splicing of *rem1* from a wild-type promoter but not in a FLEX box mutant background (Fig. 3a). Furthermore, *rem1* with mutations at the FLEX elements show induced transcription at early time points of meiosis, when there is no expression of *mei4* (Supplementary Fig. 4b, ΔFLEX2 panel). Together, these data indicate that other forkheads could also bind to these FLEX boxes and regulate transcription or splicing. Whereas deletion of *sep1* or *fhl1* had no effect, cells lacking Fkh2 showed *rem1* splicing during vegetative growth as well as meiosis (Fig. 3b, c and Supplementary Fig. 5). We determined that *rem1* splicing in Δ*fkh2* cells during vegetative growth was not due to the ectopic presence of Mei4 (data not shown). These findings prompted us to examine whether intron retention was a common feature of other genes during meiosis. We decided to use whole-genome microarray analysis to compare transcriptional profiles and intron retention of wild-type and Δ*mei4* strains. RNA was isolated from both strains at 0 h and 4 h after meiotic induction, and microarray data were collected mainly as described elsewhere[14]. Although the microarrays did not include probes for all *S. pombe* introns, the analyses showed that a set of genes presented intron retention in the absence of Mei4 (Supplementary Fig. 6a). Each one of these genes was analysed by RT–PCR, showing that they paralleled the pattern of *rem1* splicing during meiosis and vegetative growth (Supplementary Figs 6 and 7).

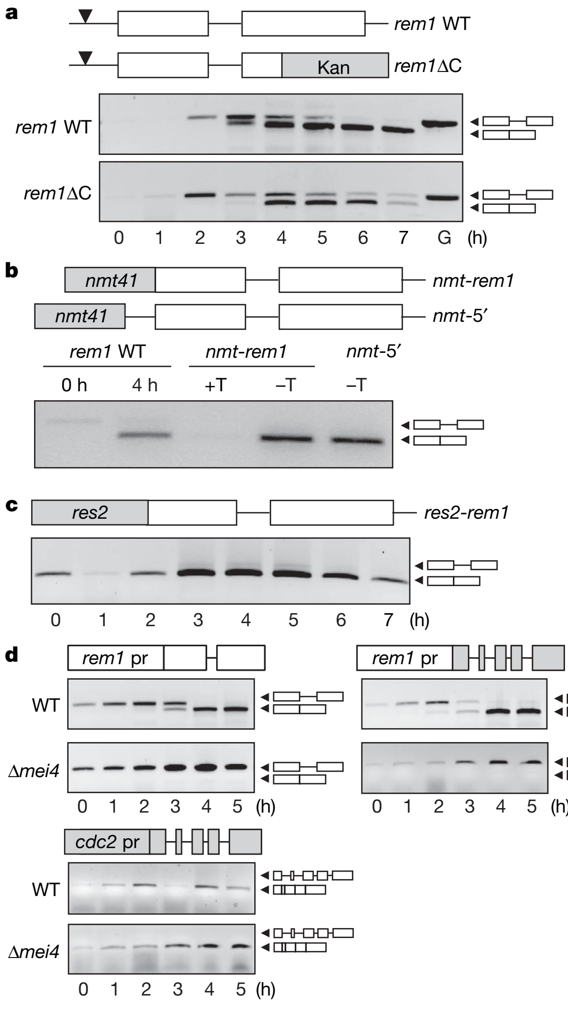


Figure 2 | The promoter of *rem1* controls splicing. **a**, *rem1* splicing analyses in a wild-type strain (*rem1* WT) and a strain deleted at the 3′ end of *rem1* (*rem1*ΔC). The hours of meiosis are indicated; G, genomic DNA. Arrowheads indicate the transcription start site (−160 nucleotides). Kan, kanamycin cassette. **b**, *rem1* splicing analyses in cells carrying *nmt-rem1* and *nmt*-5′ chimaeras (with or without *rem1* 5′ UTR). Cells were grown with (promoter off) or without (promoter on) thiamine. Wild-type strains at 0 h and 4 h of meiosis were used as controls. **c**, *rem1* splicing analyses in cells carrying *res2-rem1* chimaera during meiosis. **d**, Splicing analyses of endogenous *rem1* and *cdc2* or *rem1-cdc2* chimaera in wild-type and Δ*mei4* strains. pr, promoter.

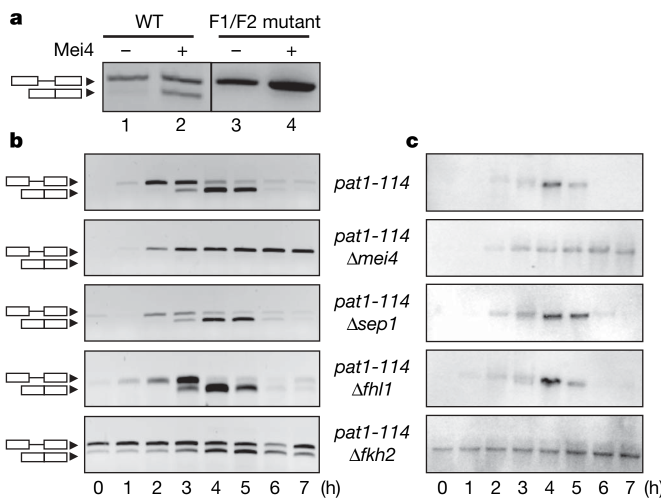


Figure 3 | *rem1* splicing is dependent on forkhead transcription factors. **a**, Overexpression of Mei4 (lanes 2 and 4) in mitotically growing cells induces splicing of wild-type *rem1* (lane 2) but is unable to induce splicing when the two FLEX boxes are mutated (lane 4; F1/F2 mutant). Fkh2 and Mei4 regulate the splicing of *rem1*. The rate of splicing (**b**) and accumulation of *rem1* mRNA (**c**) was analysed by RT–PCR and northern blot in the indicated strains. The hours of meiosis are indicated at the bottom.

To investigate further the role of the promoter in the regulation of splicing, we examined Mei4 and Fkh2 binding to specific promoters during meiosis and the timing of these interactions. Chromatin immunoprecipitation (ChIP) analysis showed that Mei4 and Fkh2 associated specifically with the *rem1* promoter (Fig. 4a). Notably, whereas the increased binding of Fkh2 was smoother during the time course—probably because Fkh2 was already bound to the promoter at the beginning of the meiosis (see the effect of *fkh2* deletion on splicing at early time points; Fig. 3b, lower panel)—there was an abrupt increase in binding of Mei4. Also, whereas Fkh2 was only detected in the promoter region, Mei4 was detected along the entire transcriptional unit. Next, we determined loading of RNA polymerase (Pol) II on *rem1*. ChIP analysis showed that Pol II is more efficiently loaded on the *rem1* promoter in the presence of Mei4 (Fig. 4b and Supplementary Fig. 8), although loading and transcription do occur to a certain extent in the absence of Mei4 at later time points of meiosis (see northern blots in Fig. 3c). We also noticed that in the absence of Mei4 there was an accumulation of the phospho-Ser 5 isoform of Pol II (concomitant with undetectable levels of the phospho-Ser 2 isoform), indicating that there might be a problem with Pol II elongation in Δ*mei4* cells that could have consequences on the coupling of transcription with pre-mRNA processing[15]. Next, we wanted to determine whether Mei4 and/or Fkh2 physically associated with proteins of the spliceosome to localize its function to specific genes. We tested the co-immunoprecipitation of Mei4 and Fkh2 with U1-70K (one of the 16 proteins of the U1 small nuclear ribonucleoprotein particles (snRNP) in fission yeast[16]), with Prp11 (the ATP-dependent helicase that bridges U1 and U2 snRNP, which is required for complex B formation during splicing[17] and is known as PRP5 in other organisms) and with Cdc5 (a component of the nineteen complex (NTC), which is part of the active form of the spliceosome[18]; the human homologue has been found associated with core Sm proteins[19]).

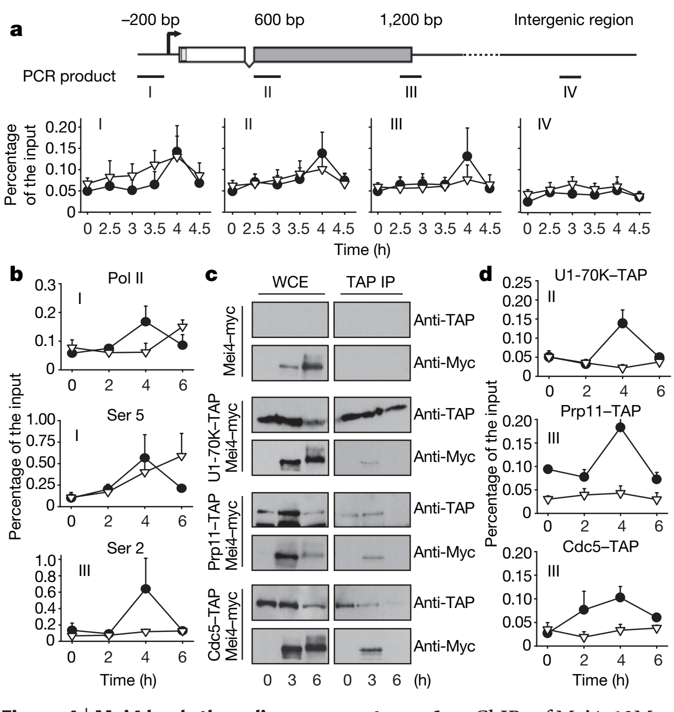


Figure 4 | Mei4 loads the spliceosome onto *rem1*. a, ChIPs of Mei4–13Myc (circles) and Fkh2–HA (triangles). *rem1* and the quantitative PCR products (I to IV) are shown at the top. **b**, ChIPs of Pol II, phosphor-Ser 5 or phosphor-Ser 2 from wild-type (circles) and Δ*mei4* (triangles) strains. **c**, Meiotic extracts from strains expressing Mei4–13Myc alone or co-expressing U1-70K–TAP, Prp11–TAP or Cdc5–TAP were IgG-sepharose-purified and proteins were detected by western blotting. TAP IP, tandem affinity purification; WCE, whole cell extracts. **d**, ChIPs of U1-70K–TAP, Prp11–TAP or Cdc5–TAP from wild-type (circles) and Δ*mei4* (triangles) strains. All the ChIPs were assayed for the presence of *rem1* at the indicated meiotic time points and the mean and s.e.m. were plotted ($n = 3$).

As shown in Fig. 4c and Supplementary Fig. 9, Mei4, but not Fkh2, was detected in the immunoprecipitates of U1-70K, Prp11 and Cdc5 at the time of *rem1* splicing. The interaction between Mei4 and the active spliceosome could explain why Mei4 is required for the regulated splicing of several meiotic genes. We next determined whether different splicing factors are recruited to specific genes in a Mei4-dependent manner. We could detect U1-70K, Prp11 and Cdc5 bound to the *rem1* open reading frame, with enrichment around the intron (Fig. 4d and Supplementary Fig. 10) and with maximum binding occurring concomitantly with the onset of meiosis I and *rem1* splicing. However, in the absence of Mei4, the binding of U1-70K, Prp11 or Cdc5 was abolished. Similar results were observed with two other genes, *spo4* and *mug137*, the splicing regulation of which paralleled the pattern of *rem1* (data not shown). Thus, loading of the active spliceosome onto several meiotic genes and its splicing depends on the presence of Mei4.

We report here that *rem1*, when it is not spliced, translates into a truncated protein that brings about normal levels of recombination. This could account for a sophisticated alternative splicing scheme in a unicellular organism which will generate a 17 kDa protein with a role in recombination when the mRNA is not processed (during the pre-meiotic S phase), and alternatively will produce a protein with a described function as a cyclin during meiosis I when the mRNA is processed (ref. 1 and Supplementary Fig. 11). This occurs as part of a regulatory mechanism that ensures the absence of cyclin in mitotically growing cells, because even very low levels of Rem1 are toxic in non-meiotic cells[1]. Other mechanisms have been described to ensure the absence of meiotic-specific transcripts during vegetative growth[20]. We propose here that the regulation of *rem1* splicing is exclusively under the control of its own promoter. In this model, the Mei4 transcription factor recruits the active spliceosome to specific genes, and the Fkh2 transcription factor interferes with coupling splicing to transcription. In budding yeast, it has been shown that two forkhead proteins, Fkh1 and Fkh2, have opposing functions in regulating Pol II elongation[10]. Thus, a switch from Fkh2 to Mei4 could explain the effect on splicing regulation. Other possibilities could be that Fkh2 and/or Mei4 act as chromatin-remodelling factors on specific promoters[21], interfering with the coupling of splicing to transcription, or they could be affecting the processivity of Pol II, as in mammalian cells in which evidence for coupling between alternative splicing and promoter-specific transcription has been shown[22].

METHODS SUMMARY

Strains and media. All *S. pombe* strains used are isogenic to the wild-type 972h$^-$ strain and are described fully in the Supplementary Information. Media was prepared as described[23]. *S. pombe* cells were transformed using the lithium acetate method as described[23].

Synchronous meiosis. To obtain synchronous meiotic cultures in the *pat1-114* background, we used standard procedures described elsewhere[1,24]. Synchronicity was assessed by flow cytometric analysis and 4′,6′-diamidino-2-phenylindole (DAPI) staining as described previously[25].

Gene expression analysis. RNA was prepared by lysis in the presence of hot phenol, as described[26]. For the RT–PCR reactions, 8 μg of DNase-treated RNA was denatured at 65 °C for 10 min and then chilled on ice. Reverse transcription reactions were carried out with oligodT priming following manufacturer's guidelines (Promega), in the presence or absence of the enzyme. One microlitre of the newly synthesized complementary DNA was used in the PCR reactions with specific primers[1]. Global expression analysis was carried out as described in the Supplementary Methods.

Protein extraction and immunoprecipitations. Extracts were prepared in NET-N buffer[1] and immunoprecipitations were performed with 30 μl of immunoglobulin G (IgG) sepharose. Immunoprecipitates (8 mg) were performed in IPP buffer (10 mM Tris, pH 8.0, 150 mM NaCl, 0.1% NP-40, 1 mM dithiothreitol, 0.5 mM phenylmethyl sulphonyl fluoride (PMSF), 5 μg ml^{-1} leupeptin and 5 μg ml^{-1} aprotinin), washed three times with the same buffer, resolved in 8% SDS–PAGE, transferred to nitrocellulose membranes and blotted with the indicated antibody.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 May; accepted 6 August 2008.
Published online 24 September 2008.

1. Malapeira, J. *et al.* A meiosis-specific cyclin regulated by splicing is required for proper progression through meiosis. *Mol. Cell. Biol.* **25,** 6330–6337 (2005).
2. Horie, S. *et al.* The *Schizosaccharomyces pombe mei4*$^{+}$ gene encodes a meiosis-specific transcription factor containing a forkhead DNA-binding domain. *Mol. Cell. Biol.* **18,** 2118–2129 (1998).
3. Bulmer, R. *et al.* The forkhead transcription factor Fkh2 regulates the cell division cycle of *Schizosaccharomyces pombe. Eukaryot. Cell* **3,** 944–954 (2004).
4. Nielsen, O. & Egel, R. The pat1 protein kinase controls transcription of the mating-type genes in fission yeast. *EMBO J.* **9,** 1401–1406 (1990).
5. Mata, J., Lyne, R., Burns, G. & Bahler, J. The transcriptional program of meiosis and sporulation in fission yeast. *Nature Genet.* **32,** 143–147 (2002).
6. Primig, M. *et al.* The core meiotic transcriptome in budding yeasts. *Nature Genet.* **26,** 415–423 (2000).
7. Chu, S. *et al.* The transcriptional program of sporulation in budding yeast. *Science* **282,** 699–705 (1998).
8. Averbeck, N., Sunder, S., Sample, N., Wise, J. A. & Leatherwood, J. Negative control contributes to an extensive program of meiotic splicing in fission yeast. *Mol. Cell* **18,** 491–498 (2005).
9. Carlsson, P. & Mahlapuu, M. Forkhead transcription factors: key players in development and metabolism. *Dev. Biol.* **250,** 1–23 (2002).
10. Morillon, A., O'Sullivan, J., Azad, A., Proudfoot, N. & Mellor, J. Regulation of elongating RNA polymerase II by forkhead transcription factors in yeast. *Science* **300,** 492–495 (2003).
11. Ribar, B., Banrevi, A. & Sipiczki, M. *sep1*$^{+}$ encodes a transcription-factor homologue of the HNF-3/forkhead DNA-binding-domain family in *Schizosaccharomyces pombe. Gene* **202,** 1–5 (1997).
12. Szilagyi, Z., Batta, G., Enczi, K. & Sipiczki, M. Characterisation of two novel fork-head gene homologues of *Schizosaccharomyces pombe*: their involvement in cell cycle and sexual differentiation. *Gene* **348,** 101–109 (2005).
13. Buck, V. *et al.* Fkh2p and Sep1p regulate mitotic gene transcription in fission yeast. *J. Cell Sci.* **117,** 5623–5632 (2004).
14. Bozdech, Z. *et al.* Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray. *Genome Biol.* **4,** R9 (2003).
15. Komarnitsky, P., Cho, E. J. & Buratowski, S. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.* **14,** 2452–2460 (2000).
16. Newo, A. N. *et al.* Proteomic analysis of the U1 snRNP of *Schizosaccharomyces pombe* reveals three essential organism-specific proteins. *Nucleic Acids Res.* **35,** 1391–1401 (2007).
17. Huang, T., Vilardell, J. & Query, C. C. Pre-spliceosome formation in *S.pombe* requires a stable complex of SF1–U2AF^{59}–U2AF^{23}. *EMBO J.* **21,** 5516–5526 (2002).
18. Ohi, M. D. *et al.* Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel pre-mRNA splicing factors, and snRNAs. *Mol. Cell. Biol.* **22,** 2011–2024 (2002).
19. Burns, C. G., Ohi, R., Krainer, A. R. & Gould, K. L. Evidence that Myb-related CDC5 proteins are required for pre-mRNA splicing. *Proc. Natl Acad. Sci. USA* **96,** 13789–13794 (1999).
20. Harigaya, Y. *et al.* Selective elimination of messenger RNA prevents an incidence of untimely meiosis. *Nature* **442,** 45–50 (2006).
21. Hatta, M. & Cirillo, L. A. Chromatin opening and stable perturbation of core histone: DNA contacts by FoxO1. *J. Biol. Chem.* **282,** 35583–35593 (2007).
22. Cramer, P., Pesce, C. G., Baralle, F. E. & Kornblihtt, A. R. Functional association between promoter structure and transcript alternative splicing. *Proc. Natl Acad. Sci. USA* **94,** 11456–11460 (1997).
23. Moreno, S., Klar, A. & Nurse, P. Molecular genetic analysis of the fission yeast *Schizosaccharomyces pombe. Methods Enzymol.* **194,** 795–823 (1991).
24. Murakami, H. & Nurse, P. Meiotic DNA replication checkpoint control in fission yeast. *Genes Dev.* **13,** 2581–2593 (1999).
25. Ayte, J., Schweitzer, C., Zarzov, P., Nurse, P. & DeCaprio, J. A. Feedback regulation of the MBF transcription factor by cyclin Cig2. *Nature Cell Biol.* **3,** 1043–1050 (2001).
26. Castillo, E. A. *et al.* Diethylmaleate activates the transcription factor Pap1 by covalent modification of critical cysteine residues. *Mol. Microbiol.* **45,** 243–254 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Nurse, K. Gould, C. Query, N. Kaufer, H. Murakami and C. McInerny for strains and plasmids; J. Wu and Y. Zhurinskiy for help and advice on ChIPs; E. Hidalgo, J. Valcárcel, J. Vilardell and G. Gil for critical reading of the manuscript; and members of the Oxidative Stress and Cell Cycle Group at Universitat Pompeu Fabra and P. Nurse's laboratory at Rockefeller University for suggestions and comments. We acknowledge the technical support of M. Carmona and K. Zueckert-Gaudenz. J.A. is supported by a contract Ramon y Cajal from the Ministerio de Educación y Ciencia. This work was supported by grants from the Ministerio de Ciencia y Tecnologia (BFU2006-01785) and the Consolider-Ingenio 2007-0020.

Author Information Microarray data are available at ArrayExpress (www.ebi.ac.uk/array express/) under the accession number E-TABM-465. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.A. (jose.ayte@upf.edu).

METHODS

Growth and maintenance of *S. pombe* strains. All *S. pombe* strains used are isogenic to wild-type 972 h^- and are listed in Supplementary Table 1. Media were prepared as described previously[27] and *S. pombe* was transformed using the lithium acetate method[28]. Flow cytometry was performed as described previously[29]. For DAPI staining, 1 ml of cell culture (1×10^7 to 2×10^7 cells) was briefly centrifuged, fixed with 70% ethanol and stored. Fifty microlitres of fixed cells ($\sim 5 \times 10^5$ cells) were centrifuged and resuspended with 25 µl of water, spotted onto slides and fixed by heating. Nuclei were stained with DAPI (5 µg ml^{-1}). Recombination efficiency was determined as described[30]. In brief, intragenic *ade6* recombination (gene conversion) was measured in wild-type and Δ*rem1* strains with the *ade6* alleles *ade6-M26* and *ade6-52*. Ade$^+$ prototrophs per 10^6 viable spores were measured and the efficiency relative to a wild-type strain (expressed as a percentage) was plotted.

Synchronous meiosis was performed as described previously[30] except for the experiments that involved co-immunoprecipitations (CoIPs), which were performed as follows: single colonies were grown in YE5S (ref. 27) to early stationary phase. The cells were diluted in minimal medium supplemented with 100 µg ml^{-1} leucine and grown at 25 °C to 1×10^7 to 2×10^7 cells ml^{-1}. The culture was then shifted to 35.5 °C in a water shaker (INFORS HT) to induce meiosis.

Gene expression analysis. Each cell pellet was resuspended in 0.4 ml of AE buffer (50 mM sodium acetate, pH 5.3, 10 mM EDTA, pH 8.0). SDS was then added to a final concentration of 1%, and proteins and DNA were extracted by adding 0.6 ml of 1:1 acidic phenol:chloroform and incubation at 65 °C for 1 h. The aqueous phase was separated by centrifugation at 10,000*g* for 2 min at 4 °C. The sample was washed twice with 0.2 ml of 1:1 acidic phenol:chloroform. After chloroform extraction, RNA was precipitated with ethanol. RNA concentration was determined by absorbance at 260 nm, and equal amounts (5 µg) were loaded in formaldehyde-agarose gels[30] containing ethidium bromide to confirm equal loading by visualizing ribosomal RNA. RNA was then transferred to GeneScreen Plus nylon membranes (PerkinElmer Life Sciences). Hybridization and washes were performed as recommended by the manufacturer.

Splicing was analysed by RT–PCR. RNA was digested with DNase I for 30 min at 37 °C, phenol extracted and precipitated. Eight micrograms of total RNA was denatured at 65 °C for 10 min and then chilled on ice. Reverse transcriptase reactions were carried out (60 min at 42 °C, 30 min at 52 °C and 3 min at 94 °C) following the manufacturer's guidelines (Promega) in the presence or absence of the enzyme. One microlitre of the cDNA was used in the PCR reactions with the primers listed in Supplementary Table 3. Genomic DNA was used in Fig. 2a as a control of unspliced *rem1*.

Microarray experiments and data evaluation. Global expression analysis used custom designed *S. pombe* microarrays. Array construction, sample labelling and hybridization were carried out as described previously[31]. The arrays consisted of 8,785 70-mer oligonucleotides representing 6,918 exons, 1,521 introns and 333 3′ UTRs (http://research.stowers-institute.org/microarray/S_pombe/). In brief, total RNA was prepared from wild-type and Δ*mei4* strains, at 0 h and 4 h from meiosis induction. Polyadenylated RNA was extracted from total RNA by purification with an oligo(dT) cellulose column. RNA quality was assessed on a Bioanalyser 2100 machine (Agilent). RNA (2 µg per sample) was converted to cDNA by priming with oligo-dT18 and poly-dN9 in the presence of aminoallyl-dUTP (Ambion), followed by conjugation to Cy5 or Cy3 fluorescent dyes. Samples were dye-swapped for further technical replication. Labelled samples were mixed for comparative hybridization on poly-L-lysine-coated microarrays, which were printed with 70-base oligomers representing all known *S. pombe* reading frames and introns greater than 70 bases in length. The microarrays were scanned with a GenePix 4000B scanner and the images were analysed using GenePix Pro 6.0 software (Molecular Devices). Data analysis was performed with the R programming language. Differential expression was assessed using the Limma package[32]. Eight microarrays were used, with two experimental factors: time points 0 h and 4 h from meiotic induction and wild-type and Δ*mei4* strains. Data within arrays were normalized using the print-tip loess method. Data between arrays were normalized using Aquantile normalization. Linear models were fit with coefficients estimated for wild type 0/mei4 0, wild type 4/mei4 4, wild type 0/mei4 4 and mei4 0/mei4 4. Microarray data are available at ArrayExpress (http://www.ebi.ac.uk/array express/) under the accession number E-TABM-465.

After defining the genes that parallel the pattern of *rem1*, we have aligned the sequences of their promoters and searched for similar patterns in the direct and reverse strands. We have obtained the frequency of every nucleotide at each position (Supplementary Fig. 6b). The frequency of the nonamer FLEX box GTAAACAAA in the *S. pombe* genome is: 214 genes have 1 GTAAACAAA sequence, of which 9 of the genes have 2 GTAAACAAA sequences; 332 genes have 1 TTTGTTTAC sequence, of which 24 of the genes have 2 TTTGTTTAC sequences and 1 gene has 3 TTTGTTTAC sequences. The data were obtained by searching for FLEX boxes in both strands of the promoters of *S. pombe*.

Chromatin immunoprecipitation. Ninety millilitres of cells at an A_{600} of 0.5 (or equivalent) were crosslinked using 1% formaldehyde (final concentration) for 10 min at 25 °C. The pellet was washed twice with PBS. Cells were resuspended in breaking buffer (0.1 M Tris, pH 8.0, 20% glycerol and 1 mM PMSF) and broken in a Minibead-beater three times at medium intesity. The pellet was washed twice with lysis buffer (50 mM HEPES, pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS and 1 mM PMSF), resuspended in 0.25 ml of lysis buffer and sonicated to a mean chromatin fragment size of 500 bp on a Diagenode water sonicator. The chromatin was centrifuged for 15 min at 16,000*g* and the pellet was discarded. Immunoprecipitations were performed overnight with the same amount of protein (measured by Bradford) with 15 µl IgG-sepharose (Amersham) or a specific antibody (9E10 for Myc-tagged proteins, 12CA5 for HA-tagged proteins, anti RNA Pol II CTD monoclonal antibody (8WG16, Abcam), anti-RNA Pol II CTD (phospho-Ser5) polyclonal antibody (Abcam) or anti RNA Pol II CTD (phospho-Ser2) polyclonal antibody (Abcam)). Immunoprecipated extracts were incubated with 15 µl of protein G sepharose for 5 h. Beads were washed once with lysis buffer, twice with lysis buffer containing 500 mM NaCl, twice with wash buffer (10 mM Tris, pH 8, 0.25 M LiCl, 0.5% NP-40, 0.5% sodium deoxycholate, 1 mM EDTA and 1 mM PMSF) and once with TE buffer (10 mM Tris, pH 8, 1 mM EDTA). The beads were resuspended in 0.1 ml of elution buffer (50 mM Tris, pH 8, 10 mM EDTA, 1% SDS) and incubated at 65 °C for 20 min. The beads were then washed with 0.15 ml of TE buffer with 0.67% SDS and the crosslinking was reversed overnight at 65 °C. To clean up the DNA, 0.25 ml of TE buffer was added, 1 µl of 20 mg ml^{-1} glycogen and 7.5 µl of 20 mg ml^{-1} Proteinase K and incubated for 2 h at 37 °C. DNA was extracted with phenol/chloroform and precipitated with ethanol and sodium acetate for 30 min at −20 °C. DNA was pelleted for 30 min at 16,000*g* at 4 °C. The pellet was air-dried and resuspended in 100 µl of TE buffer. To perform the PCR, a SYBR green kit was used and quantitative real time PCR was performed in an Applied Biosystems Cycler. Data were expressed as a percentage of the input.

27. Moreno, S., Klar, A. & Nurse, P. Molecular genetic analysis of the fission yeast *Schizosaccharomyces pombe*. *Methods Enzymol.* **194**, 795–823 (1991).
28. Maundrell, K. Thiamine-repressible expression of vectors pREP and pRIP for fission yeast. *Gene* **123**, 127–130 (1993).
29. Ayte, J., Schweitzer, C., Zarzov, P., Nurse, P. & DeCaprio, J. A. Feedback regulation of the MBF transcription factor by cyclin Cig2. *Nature Cell Biol.* **3**, 1043–1050 (2001).
30. Malapeira, J. *et al.* A meiosis-specific cyclin regulated by splicing is required for proper progression through meiosis. *Mol. Cell. Biol.* **25**, 6330–6337 (2005).
31. Sinibaldi, R. *et al.* Gene expression analysis on medium-density oligonucleotide arrays. *Methods Mol. Biol.* **170**, 211–222 (2001).
32. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, Article–3 (2004).

naturejobs

THE CAREERS MAGAZINE FOR SCIENTISTS

The financial crisis that continues to reverberate worldwide promises to affect plenty of scientists. Boston University in Massachusetts is one of the first to freeze hiring and halt new building projects. And according to a 1 October article in the *Boston Globe*, the American Council on Education says that many more institutions are likely to follow suit.

A prolonged downturn might mean that the hiring of adjunct staff in place of tenure-track faculty members, a practice that is gaining popularity because it saves money, could become even more commonplace. Tuition fees are likely to rise, and as student loans get harder to secure, private colleges might be forced to chip away bigger chunks of their endowments to support students.

Some endowments have taken a sizeable hit. London's Wellcome Trust, the world's largest medical charity, has had its endowment drop by up to 10% to roughly £13 billion (US$23 billion), mostly because of stock-market gyrations. New York's Cold Spring Harbor Laboratory has seen its $300 million endowment drop by around 5% (see *Nature* **455,** 712–713; 2008). And the charitable giving that is the lifeblood of many science institutions could well drop off in the coming months.

Other organizations are likely to fare better, however. The Howard Hughes Medical Institute in Chevy Chase, Maryland, for example, whose endowment has seen minimal damage, has no current plans to cut back on projects. And many universities may see a rise in enrolment if the downturn continues for an extended period of time — dwindling job prospects often spur an exodus back to school.

Energy and environment research may still see long-term growth. In the United States, the presidential candidates routinely pledge significant investments in alternative energies and climate-change programmes, saying that these expenditures will help create millions of 'green jobs' (see *Nature* **455,** 565; 2008). Of course, many campaign promises are not realized, and the crisis is still in its early stages. Already dealing with slimmed-down budgets, researchers would do well to monitor the economic health of their institution and the field of their choice.

Gene Russo is editor of *Naturejobs*.

CONTACTS
Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street, London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
***Nature*events:**
Ghizlaine Ababou (+44 (0) 20 7014 4015)
e-mail: g.ababou@nature.com
UK Corporate:
Nils Moeller (4953)
Southwest UK/RoW:
Alexander Ranken (4944)
Northeast UK/Ireland:
Matthew Ward (+44 (0) 20 7014 4059)
France/Switzerland/Belgium:
Muriel Lestringuez (4994)
Scandinavia/Spain/Portugal/Italy:
Evelina Rubio-Hakansson (4973)
North Germany/The Netherlands/Eastern Europe: Reya Silao (4970)
South Germany/Austria:
Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
***Naturejobs* web development:** Tom Hancock
***Naturejobs* online production:** Dennis Chu

US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

India
Vikas Chawla (+91 1242881057)
e-mail: v.chawla@nature.com

Japan Head Office, Tokyo
Chiyoda Building, 2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe (+81 3 3267 8765)
e-mail: a.watanabe@natureasia.com
Business Development Manager, Greater China/Singapore:
Gloria To (+852 2811 7191)
e-mail: g.to@natureasia.com

Y.R. GAITONDE, CENTER FOR AIDS RESEARCH; B. AKERS

G. CAMPAIGN; J. ASHER

From left: Suniti Solomon, Renée Sharp, Suman Sahai and Jana Asher.

MAKING A DIFFERENCE

The demand for science skills in non-governmental organizations is rising. **Paroma Basu**, **Jane Qiu** and **Kendall Powell** report.

In the early 1980s, Arun Gupta, a doctor working in the north Indian state of Punjab, conducted a survey on Indian women's breastfeeding habits. Of the 100 women he questioned in his clinic, all had been given a tin of infant formula, even before their babies had been born. Using infant formula instead of breastfeeding, especially in places where clean water is not guaranteed, can have detrimental effects on babies. On further investigation, Gupta found that the practice was rampant: companies had major discount deals, sometimes offering ten free tins of formula for every ten bought. And they provided free samples to nursing homes and hospitals, which then passed them on to mothers, who could have difficulty trying to breastfeed later if their babies were used to bottles. Gupta had serious concerns that women were being diverted from breastfeeding by a marketing campaign.

So he and a few colleagues decided to act. In 1991, they pooled some money and started a formal organization advocating breastfeeding. They registered this non-governmental organization (NGO) in 1992, and called it the Breastfeeding Promotion Network of India. The organization teaches health-care workers how to introduce healthy breastfeeding practices to new mothers.

"NGOs can have a very significant role in a developing country such as India," says Asish Ghosh, director of the Centre for Environment and Development in Calcutta. "In a democracy we need to have a participatory management system and processes, we at least need to get all the stakeholders' opinions. NGOs can be very useful in enabling this."

India has thousands of NGOs, several of which have objectives related to science. As the role of NGOs in developing countries expands, new opportunities have been created for scientists to help their fellow citizens and their causes by taking on mammoth challenges, ranging from energy to environmental conservation to human-rights abuses. They are a popular vehicle for scientists fighting for principles that are affected by modernization and the rapid pace of development and population growth.

Healthy environment

Numbers of NGOs are increasing in countries such as China and India. Ghosh notes that in the environment sector, NGOs started burgeoning in the mid-1980s, followed by a burst in the 1990s. However, those focused on scientific research, rather than advocacy, are still a relatively small group. Suman Sahai, director of Gene Campaign in Delhi, cites a crucial need for science-based NGOs in an increasingly knowledge-based economy. "There are now immense technological inputs into the global economy, whether with water, agriculture or intellectual property," says Sahai, who helps local communities keep control over their genetic resources by, for example, fighting against seed patents. A science background lends both insight and credibility to your cause, argues Sahai, a PhD geneticist. "Many times, activists who are not well informed may end up hurting the policy issue," she says. "My being a professor gives me a tremendous advantage. I speak with authority, I am perceived as an expert and automatically sit on every panel."

Scientists from a variety of backgrounds are motivated by any number of reasons to work for NGOs — but 'making a difference' is a common thread. This is often a challenging prospect in developing nations, where rampant growth puts pressure on environment and agriculture, and raises pressing questions about

the impact of technology. Debal Deb, head of the Centre for Interdisciplinary Studies in Calcutta, was passionate about preserving rare varieties of native rice species by distributing a number of rice varieties for free. After establishing the first non-profit seed bank in India in 1996, he attended the University of California, Berkeley, on a fellowship. He saved money, returned to West Bengal, bought some land and started an organic rice farm, where he is growing 500 native varieties. "My only desire," Deb says, "was to save all these rice varieties to preserve and conserve all this genetic diversity that is getting lost."

Research skills needed

In China, devastating pollution led Ma Jun to found the Institute of Public and Environmental Affairs (IPEA) in 2006. Formerly an investigative journalist, Ma put his research skills to good use and is now an environmental consultant. His water-pollution map of China was followed by an air-pollution map this year: both document polluted areas and the purported sources of that pollution. The maps have identified thousands of companies, including hundreds of multinationals, whose pollution exceeds Chinese limits. Companies who clean up their acts have their names removed from the list. Based in Beijing, the IPEA is working on health and environment data as well as the pollution maps of supply chains.

Investigating humans-rights issues in war-torn countries.

J. ASHER

As well as scientific knowledge and research skills, analytical expertise can be applied to good causes. Jana Asher, a statistician who consults for NGOs, has spent time in war-torn countries, monitoring and documenting human-rights abuses. Asher collects and analyses human-rights violations data, including mortality studies in Iraq and a survey of abuses after Sierra Leone's 11-year civil war. Her research involves designing new survey methods for populations, such as those displaced in Sierra Leone, for whom traditional surveys fail to collect the proper data. The work was not for the faint of heart. "You have to be prepared to get sick or be evacuated at a moment's notice," she says. "And you have to be honest with your family about what the risks are."

As in the developed world, the academic backgrounds of NGO scientists working in developing countries vary widely. Anumita Roychowdhury, associate director of research and advocacy at the Delhi-based Centre for Science and Environment (CSE), says most of her 100 colleagues "have not done anything consciously in terms of planning or training to work at an NGO". Work there, she says, combines the rigour of academic research and technical precision of science with the incisiveness of journalistic investigation, and action on policies. Scientific staff include engineers, physicians and PhDs in the sciences. The CSE also looks for passionate people. They should be a "little mad about things", says Roychowdhury.

Ghosh divides staff into two types: those fresh from university, usually with a master's degree, and more senior professionals with ties to a university or other organization. Colleagues' backgrounds range from agriculture to engineering to geology to forestry to life sciences, he says. In the Indian city of Chennai, the Y. R. Gaitonde Centre for AIDS Research encourages more specialization. It has 11 doctors, 15 or 16 nurses, plus counsellors with master's degrees in social work, MBAs, several people working on PhDs and, crucially, biostatisticians. "Without them we cannot publish papers," says founding director Suniti Solomon.

Attracting people with the right skills has been a challenge for Ma at the IPEA. He needs staff with academic training in environmental engineering, business management, data analysis and information technology, skill sets also hotly pursued by government and business. Language skills are also, not surprisingly, highly sought-after in places such as China and India.

Working for an NGO offers plenty of perks, as well as many potential drawbacks. Ideally, it means getting paid to work for a cause you are passionate about.

"It's the possibility of making a difference, and to change at least some parts of the world," says Andreas Wilkes, an expert on climate adaptation at the World Agroforestry Centre's Beijing office. Wilkes notes that in grassroots NGOs, researchers may have direct influence over a small number of people in a specific place. In a think tank, they could influence policies that affect much bigger populations. "When I came here 10 years ago, most Chinese people didn't know what NGOs were," he says. "There's much more awareness now." He lauds the rewards of "participatory research", in which the main stakeholders play a part, such as involving farmers in an agricultural study. It's aimed at solving practical problems, says Wilkes, "and the results are better assimilated by the community".

> "Your personal work isn't going to save the world. It's a bazillion groups all working together in parallel messy ways."
> — Renée Sharp

Rich rewards

NGO positions offer a sharp contrast to government jobs in developing countries. India's government, for example, is famously bureaucratic. "Compared with working in government, I have a lot of freedom," says Solomon. However, she emphasizes that her centre's HIV research activities still must be approved via clearance from the Indian Council of Medical Research, often making such tasks painstaking and arduous.

NGO work can be very consuming and, like academic research, the victories can be few and far between, cautions Renée Sharp, an ecologist and senior analyst with the Environmental Working Group in Oakland, California. She recommends taking a long-term view. "You and your personal work aren't going to save the world," says Sharp. "It's a bazillion groups all working together in parallel and messy ways." ■

Paroma Basu is based in Delhi, India; Jane Qiu in Beijing, China; and Kendall Powell in Broomfield, Colorado. All are freelance journalists.

NATURE|Vol 455|16 October 2008

MOVERS

Maria Betti, director, Marine Environment Laboratories, International Atomic Energy Agency, Monaco

1991–2008: Head, analytical chemistry sector, Institute for Transuranium Elements, Karlsruhe, Germany
1989–91: Researcher, National Research Council, Institute of Instrumental Analytical Chemistry, Pisa, Italy
1987–90: Assistant researcher, chemistry department, University of Pisa, Italy

Maria Betti might have expected to follow her family's tradition and study language, art or literature. Instead, she opted for chemistry, biology and physics, studying the effects of pollutants on human and environmental health. She would go on to develop novel techniques to detect trace pollutants in the environment. As she assumes the directorship of the Marine Environment Laboratories at the International Atomic Energy Agency (IAEA), she says her background will help her forge new studies of the marine environment.

Fascination with algal movements in the Mediterranean Sea led Betti to a degree in marine science at the University of Pisa, in her home town. She later received a PhD in chemistry there too, investigating trace metals in marine ecosystems — part of an environmental project at a national park. There, she found her niche. "I dedicated part of my research to implementing new instrumental techniques to detect inorganic pollutants at trace levels," she says.

At the Institute of Instrumental Analytical Chemistry, part of Italy's National Research Council, Betti collaborated on research in the Antarctic Zone. Then, in 1991, she moved to the Institute for Transuranium Elements (ITU), part of the European Commission's Joint Research Centre. She says that changing her focus to the detection of radionuclides in the environment was the defining moment of her career. Since then, she has monitored the clean-up of nuclear accidents as well as analysing samples collected during United Nations inspections.

Betti's laboratory was the first civil installation in the world to use mass spectrometry in the analysis of radioactive microparticles, one of the tools used today by the IAEA and the European body Euratom to detect undeclared nuclear activities. That brought her to the attention of the IAEA and led to consulting and advisory roles there, says ITU spokesperson Gabriele Tamborini.

In her new position, Betti hopes to strengthen the links between European Union research organizations and the IAEA on a host of environmental matters. She plans to promote nuclear techniques to study the ocean, climate change and sustainable development of marine resources. For example, she hopes to use radionuclides to model phytoplankton and trace-element dynamics in the ocean — an ill-defined part of the global carbon cycle.

"This position," says Betti, "is the culmination of all my scientific and human experiences over the past 25 years." ■

Virginia Gewin

NETWORKS & SUPPORT

Counting fish

The Fisheries Service of the US National Oceanic and Atmospheric Administration (NOAA) is looking for a few good fish counters. 'Stock-assessment scientists' require a background in maths, computer science and/or conservation — and they are in demand.

A US government report predicts a severe shortfall in the next decade. *The Shortage in the Number of Individuals with Post-Baccalaureate Degrees in Subjects Related to Fishery Science*, by the departments of commerce and education, says the nation will need at least 340 scientists trained to do the research required to rebuild stocks and restore marine species — but US institutions will produce only 160.

"This can be an incredibly rewarding job. It's the interface between conservation and sustainability and policy-making," says Steven Murawski, director of scientific programmes and chief science adviser for the Fisheries Service. "You have the opportunity to make a difference." Stock assessment scientists gather data on species populations, on the basis of catches and aerial surveys. The data inform mathematical models that help design monitoring programmes and predict populations under different management scenarios. This in turn helps regulators to set quotas.

The current shortfall stems from the retirement of scientists hired in the 1960s and 1970s, from new mandates created by the US Magnuson–Stevens Fishery Conservation and Management Reauthorization Act of 2006, and from a lack of funding for academics. Whereas some scientists bring in millions of federal dollars, stock-assessment scientists bring in a couple of thousand from NOAA. "It's been an easy choice [for universities]," says Jim Berkson, a unit leader for the Fisheries Service. "But we're trying to make the point that this is a growth industry and it will bring in money."

Commercial fishermen want more stock-assessment scientists on deck to forestall outdated quotas. Jim Ruhle, president of Commercial Fishermen of America, cites a 2004 butterfish quota based on 2002 data.

Larry Alade, a stock-assessment scientist and research fisheries biologist with the Fisheries Service at Woods Hole, Massachusetts, says he routinely draws on his background in computer science, biology and maths. After a master's in computer science and internships at a Mississippi fishery stock-assessment programme and at Woods Hole, Alade did a PhD in marine estuary and environmental sciences. He's now got a job he loves — contributing to sustainability. ■

Karen Kaplan

POSTDOC JOURNAL

One-man team

In the fossil record they call it punctuated equilibrium. You just blink and, wham, a species appears in a bedding plane. This is my current bedtime reading. I selected it for its research relevance, but it's quickly becoming a page turner with familiar characters whose situation I can relate to. Were the first few pioneer individuals of a species intimidated by the big ecological niche waiting to be conquered? As I move to my new lecturer position, I ponder this with some empathy.

"Whose group are you working for?" is becoming a common question when I meet new faces at coffee time. It's as if they're asking, "So what species are you?" My reply: "My own." For the first time in my career I am not starting a job as part of a research team. I'm starting alone. Of course I still have distant collaborations, but nothing nearby. My group must grow if it is to persist.

Growth needs funding, and funding calls have deadlines — 32 days to go in my case. So I'm considering some ideas and meeting the locals and their research interests, face to face. Either I go it alone with my existing collaborators, or establish a new, joint project with researchers in my new department. The latter sounds good, but have I got time? I suspect it's a rate of adaptation that few of my fossilized friends would have had to endure. ■

Jon Yearsley is now a lecturer at University College Dublin, Ireland.

High output ...High impact !!!

Mushroom Rock – Natural rock formation in the campus

The University of Hyderabad created by an Act of Parliament has, over three decades of its existence, distinguished itself as one of the best in the country as a centre for teaching, research and innovations. It has carried out cutting edge research in disciplines such as the sciences, social sciences, humanities, technology, visual arts, communication, management and continuing education. The University's mission is to promote a dynamic mindset in the field of education and research, and address the challenging tasks of the contemporary world.

At a glance

- University with Potential for Excellence–recognition by UGC
- Holistic education and integration of knowledge in the Sciences, Humanities, the Social Sciences and Performing Arts
- Latest national study reveals UoH as one of the top two institutions in the country based on high publications output in SCOPUS data base
- e-Culture, e-Governance
- Fellowship to all students
- Extramural R&D Funding (to the tune of over US $ 25 million last year)
- International Outreach with provision for faculty/student exchange and Study in India Programme
- University of Hyderabad Knowledge and Innovation Park, set up on 200 acres of land, as a Joint Venture with the Government of Andhra Pradesh, to promote R&D by national and multinational agencies
- Innovative and incentive driven approach for IPR protection
- Faculty members encouraged to become entrepreneur while retaining their jobs and facilitation through the Technology Business Incubator funded by the DST, Government of India
- A cluster of established and upcoming institutions located within the campus for cross facilitation of ideas and sharing infrastructure

University of Hyderabad

Gachibowli, Hyderabad 500 046, Andhra Pradesh, India
Phone: +91 40 2313 0000, 2313 2110
www.uohyd.ernet.in

For more information on the teaching programmes and research activities of various schools and departments / centers please visit: http://www.uohyd.ernet.in/academic/school_study.html

RW171341R

Jawaharlal Nehru University (JNU), New Delhi

JNU is a residential University with most of the faculty members and students residing in a sprawling and lush green campus of 1200 acres and focuses on **Postgraduate teaching and research in Science**. In addition to having state-of-the-art facilities for conducting experimental and theoretical research in the various Schools, JNU has recently established an Advanced Instrumentation Facility, the one of its kind in the University system in India (**www.jnu.ac.in/usicjnu/index.html**).

JNU has also been recognized as a University with "**Potential for excellence**" by the **University Grants Commission** with special funding provision. The newly appointed faculty members are provided start-up funding to establish their laboratories and research programs with recurring consumable grants being available. For details about the teaching and research programs in various Science Schools and research facilities log on to: **www.jnu.ac.in**.

Applications are invited for faculty positions at **Assistant Professor**, **Associate Professor** and **Professor Levels** in the following Science Schools of Jawaharlal Nehru University (JNU), New Delhi, India:

* **School of Life Sciences**
* **School of Environmental Sciences**
* **School of Computer and Systems Sciences**
* **School of Physical Sciences**
* **School of Information Technology**
* **School of Biotechnology**
* **Special Centre for Molecular Medicine**

Application form may be downloaded from the above website under the "Careers" section and sent along with CV to: **vc@mail.jnu.ac.in** or **recruitment@mail.jnu.ac.in**. For details of the specific teaching/research areas required by individual Schools, Deans/Chairperson of the Schools could also be contacted.

The detailed advertisement for faculty recruitment would be published soon in leading National newspapers and on the JNU web site

RW171531R

Indo-U.S.
Science & Technology
Forum

Department of
Science and Technology
Govt. of India

Indo-US Research Fellowships for Indian Researchers

In an effort to augment scientific excellence in emerging areas of science and technology, the Indo-US Science and Technology Forum (IUSSTF) in association with Science and Engineering Research Council of (SERC) of Department of Science and Technology (DST), announces the Indo-US Research Fellowships. The objective of these fellowships is to enable young researchers from India to carry out research in frontier areas of science and technology at a premier institution in USA. The fellowship will enable young Indian researchers to acquaint themselves with new scientific research methods and at the same time build strong collaborative linkages between the scientific communities of US and India.

Eligibility/ Application Requirements

Academic Qualifications

Master's Degree in Engineering, Technology or equivalent or Ph.D. in Science or Technology or equivalent or M.D. Degree in Medicine or equivalent

Applicants must provide proof of independent research work in internationally recognized academic journals

Age

Upto 40 years as on 31 December 2008

Employment

An affiliation in a recognized S&T institution/university/ college

Areas covered under the Fellowship

- Atmospheric and Earth Sciences
- Chemical Sciences
- Engineering Sciences
- Life Sciences
- Medical Sciences
- Mathematical and Computational Sciences
- Physical Sciences

Promising applications in areas other than the above areas may also be considered

Place of work

The applicant should have letter of acceptance from reputed US scientific/technological institution where the applicant will be undertaking the research work under the Fellowship

Fellowship includes

- Monthly stipend
- Return Airfare
- Preparatory allowances
- Conference allowances

Fellowship Duration

Minimum 3 months and upto 12 months

Proposal Guidelines and Format

Refer to www.indousstf.org

Application Deadline

31 December 2008

For immediate answers to your queries, please contact:

Dr. Smriti Trikha
Indo-US Science and Technology Forum
12 Hailey Road, Fulbright House, New Delhi-110 001
Phone: 011-42691700, Fax: 011-23321552
Email: fellowship@indousstf.org

www.indousstf.org

RW171530R

National Chemical Laboratory,

Pune, India

National Chemical Laboratory, Pune, India (NCL), one of the largest publicly funded research laboratories belonging to the **Council of Scientific & Industrial Research (CSIR) (www.csir.res.in)**, is looking for dynamic, highly motivated scientists in diverse disciplines of research. NCL currently has about 250 Ph.D scientists, 200 technical staff, 400 students pursuing Ph.D degree and 350 project staff. NCL's core research interests are in organic chemistry, advanced and functional materials, polymer science, chemical engineering science, reaction and process engineering, plant biochemistry, biochemical sciences, heterogeneous and homogeneous catalysis, theory and computational science. NCL is equipped with world-class facilities to conduct high quality research spanning the spectrum, from fundamental to applied industrial research.

NCL is looking for scientists in all age groups from entry level to leadership positions.

Scientists at NCL have an opportunity to perform research at the frontiers of science, seek rewarding collaborations with industry and leading academic institutions, both, within India and abroad. Scientists at NCL are also encouraged to work at the intersection of disciplines, such as:

- Bioactive Organic Molecules
- Chemical Biology
- Electro-active materials
- Green Chemical processes
- Materials derived from renewable resources
- Metabolic pathway engineering
- Molecular and hybrid materials
- NMR Spectroscopy
- Plant biotechnology
- Scientific computing across length and time scales
- Supramolecular Chemistry
- Synthetic Biology
- Bio-transformations
- Clean fuels
- Electron microscopy
- Health-care
- Membrane science
- Microbial & fermentation processes
- Molecular modeling
- Novel separation processes
- Renewable energy systems
- Study of complex fluids
- Surface science

Candidates with Ph.D degree in science or engineering and minimum two years of postdoctoral experience will be considered. Rank will be commensurate with qualifications and experience. Successful candidate must show demonstrated evidence of research leadership in any of above areas and must have strong interpersonal and communications skills.

Interested applicants should upload three pdf files, i.e., curriculum vitae, a descriptive research proposal and names of three referees through our online system at **www.ncl-india.org/apply**

For more information please visit: **www.ncl-india.org/CareersatNCL/showfile.jsp?mid=10**

RW171798R

C004_nat_16oct 13/10/08 14:14 Page 4

Indo-U.S. Science & Technology Forum

Who we are

The Indo-U.S. Science and Technology Forum (IUSSTF), established under an agreement between the Governments of India and the United States of America, is an autonomous, not for profit society that promotes and catalyzes Indo-U.S. collaborations in science, technology, engineering and biomedical research through substantive interaction among government, academia and industry.

What we do

Foster excellence by capitalizing on the scientific and technological synergy
Disseminate information and create awareness through scientific exchanges
Build linkages through networking between academia and industry
Explore new frontiers by nurturing contact between young and mid-career scientists
Pave way to sustainable interactions and establish long term relationships
Encourage public-private partnership to inculcate innovation and entrepreneurship

We support

Exciting and innovative collaborative programs cutting across disciplines and institutions

- Workshops/Symposium
- Student & Faculty Exchange
- Knowledge R&D Network Centres
- Industry-Academia Connect Programs
- Training/Advance Schools
- Travel Grants
- Public-Private Network Centres
- Special Initiatives for Strategic Partnership

We invite

Proposals which are peer reviewed both in India and USA for awards

Submission Deadlines	Award Announcements
15 February 2009	May 2009
15 June 2009	September 2009
15 October 2009	January 2010

How to contact us?

Further information available at
http://www.indousstf.org

IUSSTF

RW171529A

Indo-U.S. Science and Technology Forum: A Catalyst for Indo-U.S. Science & Technology Cooperation

Don't let your career go up in smoke, use Naturejobs to get the hot jobs.

naturejobs

NATURE|Vol 455|16 October 2008

The cheap crusade

Fight the system.

Rahul Kanakia

Everybody knows not to click on the spam link. Nowadays, the best thing that can happen is a virus. But when it's late at night, you just say to hell with it, like that time in middle school when you tremulously clutched the sides of the sink and whispered "Bloody Mary" three times into the mirror just to see if she'd really show up to flay you alive.


JACEY

And when the site pops up, you know you should close it right away. Because the NSA has profiles for guys like you: middle-aged stockbrokers with a gun in the drawer. They apply Bayesian analysis to reams of stats: number of pages viewed, order viewed, time viewed, interval between clicks and so on. And eventually they decide there's a non-negligible chance that you've been turned. That's when your name appears on a list. You probably won't be arrested. But you'll never get off that list.

Yet you keep reading, because the terrorists have studied you too. They bought your stats from the same corporations that sell it to the government. How do you think they even got your e-mail address? And they feed it to analysts trained, using petrodollar scholarships, at the same Harvard labs that the CIA recruits from. They know how to get your heart pumping

Oh, the terrorists don't know you will be the one who takes the bait. Out of 50 million of these e-mails sent every day, they'll be lucky to get one nibble. But they know that there's someone like you out there. And they keep e-mailing you, so you'll always have that link in your inbox at just the moment when you're at your weakest.

As you read, you feel an ecstatic rush, like when you beat your fist against the wall until the knuckles bleed or when you stop at the threshold of your door just before work, look back at the fridge, and decide to make yourself what you know will be only the first Bloody Mary of the day.

Except that this feeling is ten times better. This time there won't be a cold morning after, when you have to drag yourself through the pieces of your life, praying that today God will finally grant you the strength to resist temptation.

Only moments after you realize what you've done, a computer in Washington does too. And then the clock starts ticking. It will take only a few hours for an agent to contact your Internet service provider and find your name and address. Then he'll deputize a few of your neighbours and co-workers, under the militia clause of the Constitution, and tell them to look out for you. You won't know who they are. Hell, you've been deputized yourself.

Then you'll be truly stuck. They'll kill that little flower you've nurtured in your heart ever since you first realized what a sorry, pathetic bastard you are. Sometimes it was the only thing that kept you going as you sat at your desk, starting to regret that buzz and calculating when you could pop out to renew it.

Anytime you wanted to, you could make them pay. Oh, you wouldn't. You'd never do anything like that. But you could. And in some way, holding back gave you a kind of pride. No one knew the kind of strength it took. You were protecting them, through sheer force of will, every minute of every day. Still, it gave you a thrill to know what you could do.

But as you read, you realize that this is your last chance. Sure you can close this window, and report to work tomorrow, but they'll know. Wherever you go, they'll know. And if you ever stray; if you have irregularities in your tax statements or get into a bar fight, someone will whisper in the judge's ear, and there'll be no leniency for the likes of you.

This is good for some people. You've seen them on TV. They stared down the abyss, and they walked away. They got help. They recovered. Maybe that was why you really clicked on the link. You've always hoped that the right prayer would flip that switch inside of you, and you'd be better.

But maybe the switch flipped the other way. They make it so easy to get it done, right there. There's a field for your credit card and bank account info. Within hours they'll have laundered your money so tight that the CIA can't touch it. They have a text field where you can write a message. In 12 hours, they'll send it to everyone you know.

Then come the planning pages. They know you've been revising your plan for years. But there are still some helpful suggestions, if you want them. How to derail a train with an SUV (bonus points for lying in wait and gunning down the medics); how to make a poison pill from simple household items, as they know the last thing you want is to survive your apotheosis.

Hopefully, as you drive to the tracks with a copy of the train schedule taped to the dash and a bottle of whiskey on the passenger seat, you'll pull over and take your suicide concoction. It will be a victory of a sort.

Although, that message will still go out, and then they'll know that you even managed to fail at this. ■

Rahul Kanakia is a recent college graduate living in San Francisco.

FUTURES